

121/487

ACCA G A T T A C C A T T T T C T A G T A C C A C C G T T A C C A T T G T A T T T G A T		Majority
7151	ACCACATTACCATTTTCTACTCTAGTACCACCGTTACCATTTGTATTTGAT	cohl_al2.seq
7151	ACCACATTACCATTTTCTACTCTAGTACCACCGTTACCATTTATATTTGAT	a909_al2.seq
TGAGGTA T C T T C T A A T T T G A T A T C T C C T A C T G G A A T A A T G A C A G G T T T T A		Majority
7201	TGAGGTA T C T T C T A A T T T G A T A T C T C C T A C T G G A A T A A T G A C A G G T T T T A	cohl_al2.seq
7201	TGAGGTA T C T T C T A A T T T G A T A T C T C C T A C T G G A A T A A T G A C A G G T T T T A	a909_al2.seq
TGGTGAT A T T T T T A T T A G C A T C T G C T A A A T G G C G T C A A T A T C A A T G G A A		Majority
7251	TGGTGAT A T T T T T A T T A G C A T C T G C T A A A T G G C G T C A A T A T C A A T G G A A	cohl_al2.seq
7251	TGGTGAT A T T T T T A T T A G C A T C T G C T A A A T G G C G T C A A T A T C A A T G G A A	a909_al2.seq
T C A T A T G G G T T A T A A A T T T T A C C A T T G T A C C A C C A G C C A C G G A A A C G A T A		Majority
7301	T C A T A T G G G T T A T A A A T T T T A C C A T T G T A C C A C C A G C C A C G G A A A C G A T A	cohl_al2.seq
7301	T C A T A T G G G T T A T A A A T T T T A C C A T T G T A C C A C C A G C C A C G G A A A C G A T A	a909_al2.seq
G C C A T C T G G C A T T G T C G G A C G T C T C A G T A A G G C T G A A T G G G A G C C A T C G T		Majority
7351	G C C A T C T G G C A T T G T C G G A C G T C T C A G T A A G G C T G A A T G G G A G C C A T C G T	cohl_al2.seq
7351	G C C A T C T G G C A T T G T C G G A C G T C T C A G T A A G G C T G A A T G G G A G C C A T C G T	a909_al2.seq
C A T A G G A A T C A G G G T C A G T A G G C T C A T T A C T T G T T T G T A A T T G C T G A C C A		Majority
7401	C A T A G G A A T C A G G G T C A G T A G G C T C A T T A C T T G T T T G T A A T T G C T G A C C A	cohl_al2.seq
7401	C A T A G G A A T C A G G G T C A G T A G G C T C A T T A C T T G T T T G T A A T T G C T G A C C A	a909_al2.seq
G A A G C A T C C A A T G C T G G C T T T C C A T C T G T A C C A A C A G C A T C A T T G C T G T A		Majority
7451	G A A G C A T C C A A T G C T G G C T T T C C A T C T G T A C C A A C A G C A T C A T T G C T G T A	cohl_al2.seq
7451	G A A G C A T C C A A T G C T G G C T T T C C A T C T G T A C C A A C A G C A T C A T T G C T G T A	a909_al2.seq
T A T A A T A T G A T A A T C T C C A G C C T T T C G C C A A A T A G C T C T T A A A T T G A T A T		Majority
7501	T A T A A T A T G A T A A T C T C C A G C C T T T C G C C A A A T A G C T C T T A A A T T G A T A T	cohl_al2.seq
7501	T A T A A T A T G A T A A T C T C C A G C C T T T C G C C A A A T A G C T C T T A A A T T G A T A T	a909_al2.seq
C T T G A G T T A C A G C A C C T G A A A A G T T A T A A G G T C T A A T A C T A C C A T C T G G A		Majority
7551	C T T G A G T T A C A G C A C C T G A A A A G T T A T A A G G T C T A A T A C T A C C A T C T G G A	cohl_al2.seq
7551	C T T G A G T T A C A G C A C C T G A A A A G T T A T A A G G T C T A A T A C T A C C A T C T G G A	a909_al2.seq
T T A A C A T A A T A C C A A C C G A C T A A T T T G T A A G C G T C T T T T A C G T A C T T G T A		Majority
7601	T T A A C A T A A T A C C A A C C G A C T A A T T T G T A A G C G T C T T T T A C G T A C T T G T A	cohl_al2.seq
7601	T T A A C A T A A T A C C A A C C G A C T A A T T T G T A A G C G T C T T T T A C G T A C T T G T A	a909_al2.seq
C T T A G T T G T T G T A T C A A C A T T T C A G A G A C T A G T A T C T G T C G T A T A A T A G G		Majority
7651	C T T A G T T G T T G T A T C A A C A T T T C A G A G A C T A G T A T C T G T C G T A T A A T A G G	cohl_al2.seq
7651	C T T A G T T G T T G T A T C A A C A T T T C A G A G A C T A G T A T C T G T C G T A T A A T A G G	a909_al2.seq
C A T C T T T A G T T G A G T C G G G A T C T T T A T C T C G T G A A T C A T A C T T A T A A T A A		Majority
7701	C A T C T T T A G T T G A G T C G G G A T C T T T A T C T C G T G A A T C A T A C T T A T A A T A A	cohl_al2.seq
7701	C A T C T T T A G T T G A G T C G G G A T C T T T A T C T C G T G A A T C A T A C T T A T A A T A A	a909_al2.seq
T A T G T A C C T G A A G C A T C T T T G C A T A T A A T C C C T T G T A A T A T C T G T A T A A T C		Majority
7751	T A T G T A C C T G A A G C A T C T T T G C A T A T A A T C C C T T G T A A T A T C T G T A T A A T C	cohl_al2.seq
7751	T A T G T A C C T G A A G C A T C T T T G C A T A T A A T C C C T T G T A A T A T C T G T A T A A T C	a909_al2.seq

FIGURE 21K

122/487

```

CGGAAATACGATCACCATAATGCAAATCTAAATAGGTATCATCTGTTTTTC Majority
7801 CGGAAATACGATCACCATAATGCAAATCTAAATAGGTATCATCTGTTTTTC cohl_al2.seq
7801 CGGAAATACGATCACCATAATGCAAAGTCTAAATAGGTATCATCTGTTTTTC a909_al2.seq

ATAATTGGCCTCCGTTTGGATCAATATTGACACCATATGTTACCTTTTGC Majority
7860 ATAATTGGCCTCCGTTTGGATCAATATTGACACCATATGTTACCTTTTGC cohl_al2.seq
7851 ATAATTGGCCTCCGTTTGGATCAATATTGACACCATATGTTACCTTTTGC a909_al2.seq

CAACCTGCGATAGACTTTTAACATCATGAGGAGGCATAGTCCGTGTTAAAGTGC Majority
7910 CAACCTGCGATAGACTTTTAACATCATGAGGAGGCATAGTCCGTGTTAAAGTGC cohl_al2.seq
7901 CAACCTGCGATAGACTTTTAACATCATGAGGAGGCATAGTCCGTGTTAAAGTGC a909_al2.seq

AAATACTTGTGTTTTGTGCTTGGTCTTTTATACCATTTACCATCCCAAACAT Majority
7960 AAATACTTGTGTTTTGTGCTTGGTCTTTTATACCATTTACCATCCCAAACAT cohl_al2.seq
7951 AAATACTTGTGTTTTGTGCTTGGTCTTTTATACCATTTACCATCCCAAACAT a909_al2.seq

ACCTTGGTCCGACTAGGTTTTAGGTTGAACCGTTGTCGTATCGGGGGGCATAA Majority
8010 ACCTTGGTCCGACTAGGTTTTAGGTTGAACCGTTGTCGTATCGGGGGGCATAA cohl_al2.seq
8001 ACCTTGGTCCGACTAGGTTTTAGGTTGAACCGTTGTCGTATCGGGGGGCATAA a909_al2.seq

GAGGACAAATTTTGTCTCATATAGAACATCCCTTTACTGGAAAATTAGGAAG Majority
8060 GAGGACAAATTTTGTCTCATATAGAACATCCCTTTACTGGAAAATTAGGAAG cohl_al2.seq
8051 GAGGACAAATTTTGTCTCATATAGAACATCCCTTTACTGGAAAATTAGGAAG a909_al2.seq

CTCTGTATTATCAAGCGGATCTAAATATTTAATCTTGTATGAATTACGTT Majority
8110 CTCTGTATTATCAAGCGGATCTAAATATTTAATCTTGTATGAATTACGTT cohl_al2.seq
8101 CTCTGTATTATCAAGCGGATCTAAATATTTAATCTTGTATGAATTACGTT a909_al2.seq

CATACCATAACCACTAAGTTCAAATAAATCTTTGTCGTAGTCTCCATATTTA Majority
8160 CATACCATAACCACTAAGTTCAAATAAATCTTTGTCGTAGTCTCCATATTTA cohl_al2.seq
8151 CATACCATAACCACTAAGTTCAAATAAATCTTTGTCGTAGTCTCCATATTTA a909_al2.seq

TCGTAGTATTTCATCTCGGATTGGCACCTTTTGTTTTTGCACTCTGTTTGTCT Majority
8210 TCGTAGTATTTCATCTCGGATTGGCACCTTTTGTTTTTGCACTCTGTTTGTCT cohl_al2.seq
8201 TCGTAGTATTTCATCTCGGATTGGCACCTTTTGTTTTTGCACTCTGTTTGTCT a909_al2.seq

TGGGTTCTGATCAAATAAGGTAATTATCTGGATATAAGCTTTTGATAGTATT Majority
8260 TGGGTTCTGATCAAATAAGGTAATTATCTGGATATAAGCTTTTGATAGTATT cohl_al2.seq
8251 TGGGTTCTGATCAAATAAGGTAATTATCTGGATATAAGCTTTTGATAGTATT a909_al2.seq

TAACATTAAATCCTAGGTATTTTTCTGTAAAGGTAATTCCTCTGCTCCA Majority
8310 TAACATTAAATCCTAGGTATTTTTCTGTAAAGGTAATTCCTCTGCTCCA cohl_al2.seq
8301 TAACATTAAATCCTAGGTATTTTTCTGTAAAGGTAATTCCTCTGCTCCA a909_al2.seq

GCACCTCCCTCTGTCTGCTAAAGACTATTTGCCATCTAGTCCCTTGTTC Majority
8360 GCACCTCCCTCTGTCTGCTAAAGACTATTTGCCATCTAGTCCCTTGTTC cohl_al2.seq
8351 GCACCTCCCTCTGTCTGCTAAAGACTATTTGCCATCTAGTCCCTTGTTC a909_al2.seq

GTAGAACGGATAATTTTGAATTCTCTTCCCTTTTGGATAGAGTTTTATT Majority
8410 GTAGAACGGATAATTTTGAATTCTCTTCCCTTTTGGATAGAGTTTTATT cohl_al2.seq
8401 GTAGAACGGATAATTTTGAATTCTCTTCCCTTTTGGATAGAGTTTTATT a909_al2.seq

```

FIGURE 21L





124/487

		G T C C A A G C A G C C T T G G T G T T A A C A C A G A C C G A G G T G C A A T A T A G G T A A C Majority									
		9110		9120		9130		9140		9150	
9089		G T C C A A G C A G G T T G G C T C T G T T A A C A C A C G A C G A G G T G C A A T A T A G G T A A C coh1_al2.seq									
9086		G T C C A A G C A G C C T T G G T C T G T T A A C A C A C G A C G A G G T G C A A T A T A G G T A A C a909_al2.seq									
		T C C T G T T G A T A A G T C G C C T C T G T T G A A T T C A A C A C C G T A C G A T T C T T T A A Majority									
		9160		9170		9180		9190		9200	
9139		T C C T G T T G A T A A G T C G C C T C T G T T G A A T T C A A C A C C G T A C G A T T C T T T A A coh1_al2.seq									
9136		T C C T G T T G A T A A G T C G C C T C T G T T G A A T T C A A C A C C G T A C G A T T C T T T A A a909_al2.seq									
		A G A C A G G A T A A A G A C T T A T G T C T G A A G A T A C A G G T A A T G C T T G A A T T T C T Majority									
		9210		9220		9230		9240		9250	
9189		A G A C A G G A T A A A G A C T T A T G T C T G A A G A T A C A G G T A A T G C T T G A A T T T C T coh1_al2.seq									
9186		A G A C A G G A T A A A G A C T T A T G T C T G A A G A T A C A G G T A A T G C T T G A A T T T C T a909_al2.seq									
		G A T T C A G A A A G T G G A G C T C C A T T T T G A G T T T T A G A C C A A C C T A C A A A T A A Majority									
		9260		9270		9280		9290		9300	
9239		G A T T C A G A A A G T G G A G C T C C A T T T T G A G T T T T A G A C C A A C C T A C A A A T A A coh1_al2.seq									
9236		G A T T C A G A A A G T G G A G C T C C A T T T T G A G T T T T A G A C C A A C C T A C A A A T A A a909_al2.seq									
		T A A T G T T G A A G T G G G T G G T G A A A C T T T A A A T G A A C T T A G A T C C G C G G T T C Majority									
		9310		9320		9330		9340		9350	
9289		T A A T G T T G A A G T G G G T G G T G A A A C T T T A A A T G A A C T T A A T C C G C G G T T C coh1_al2.seq									
9286		T A A T G T T G A A G T G G G T G G T G A A A C T T T A A A T G A A C T T A G A T C C G C G G T T C a909_al2.seq									
		C G T T T C C T G A A C T A T C C A A T G G T A C T T G A T A A C G C T C C A A A A T A C T T T T A Majority									
		9360		9370		9380		9390		9400	
9339		C G T T T C C T G A A C T A T C C A A T G G T A C T T G A T A A C G C T C C A A A A T A C T T T T A coh1_al2.seq									
9336		C G T T T C C T G A A C T A T C C A A T G G T A C T T G A T A A C G C T C C A A A A T A C T T T T A a909_al2.seq									
		C C A G C A G A A T C A T C A T A T A A A G T T A C T G T C G C C A C T T T A C T C A T A A A T G C Majority									
		9410		9420		9430		9440		9450	
9389		C C A G C A G A A T C A T C A T A T A A A G T T A C T G T C G C C A C T T T A C T C A T A A A T G C coh1_al2.seq									
9386		C C A G C A G A A T C A T C A T A T A A A G T T A C T G T C G C C A C T T T A C T C A T A A A T G C a909_al2.seq									
		A C G A A C A T A A A T T T C T T T T G T C T C A G T T A C A G T T A T T G G C T C A C C A A A T T Majority									
		9460		9470		9480		9490		9500	
9439		A C G A A C A T A A A T T T C T T T T G T C T C A G T T A C A G T T A T T G G C T C A C C A A A T T coh1_al2.seq									
9436		A C G A A C A T A A A T T T C T T T T G T C T C A G T T A C A G T T A T T G G C T C A C C A A A T T a909_al2.seq									
		T A A C A G G G T C A C C A T A C T T T C C A G T A G T A G G A T C A T A G G T A T A C C A A C C A Majority									
		9510		9520		9530		9540		9550	
9489		T A A C A G G G T C A C C A T A C T T T C C A G T A G T A G G A T C A T A G G T A T A C C A A C C A coh1_al2.seq									
9486		T A A C A G G G T C A C C A T A C T T T C C A G T A G T A G G A T C A T A G G T A T A C C A A C C A a909_al2.seq									
		T T A A A A T G C T C T C C T G C T T T A A T C G T C G G A A T C C C A A C T T C T C C T A G A G A Majority									
		9560		9570		9580		9590		9600	
9539		T T A A A A T G C T C T C C T G C T T T A A T C G T C G G A A T C C G A A C T T C T C C T A G A G A coh1_al2.seq									
9536		T T A A A A T G C T C T C C T G C T T T A A T C G T C G G A A T C C C A A C T T C T C C T A G A G A a909_al2.seq									
		T T C T C C A T C T T T T A T A A T T T G A T G A T G A A C T T G C A T A C C T G A A G C T G T C A Majority									
		9610		9620		9630		9640		9650	
9589		T T C T C C A T C T T T T A T A A T T T G A T G A T G A A C T T G C A T A C C T G A A G C T G T C A coh1_al2.seq									
9586		T T C T C C A T C T T T T A T A A T T T G A T G A T G A A C T T G C A T A C C T G A A G C T G T C A a909_al2.seq									
		G G A A A T T A T A A T C A G T T C C G T C A T T A T T T T G A A A A T G G T A A G T T A A C C T A Majority									
		9660		9670		9680		9690		9700	
9639		G G A A A T T A T A A T C A G T T C C G T C A T T A T T T T G A A A A T G G T A A G T T A A C C T A coh1_al2.seq									
9636		G G A A A T T A T A A T C A G T T C C G T C A T T A T T T T G A A A A T G G T A A G T T A A C C T A a909_al2.seq									
		G G A A C T T C T G T A T T A T C C T C T T G A A C A A T T G C A T A A A T G G A G A A T G A A T C Majority									
		9710		9720		9730		9740		9750	
9689		G G A A C T T C T G T A T T A T C C T C T T G A A C A A T T G C A T A A A T G G A G A A T G A A T C coh1_al2.seq									
9686		G G A A C T T C T G T A T T A T C C T C T T G A A C A A T T G C A T A A A T G G A G A A T G A A T C a909_al2.seq									

FIGURE 21N

125/487

	TCTTTTAAAGCAACATCACTGCTAGTCTTCTAGTTTCTGCAGTATCTT	Majority
	9760 9770 9780 9790 9800	
9739	TCTTTTAAAGCAACATCACTGCTAGTCTTCTAGTTTCTGCAGTATCTT	cohl_a12.seq
9736	TCTTTTAAAGCAACATCACTGCTAGTCTTCTAGTTTCTGCAGTATCTT	a909_a12.seq
	TAGATTTTAAACTTCTGTTTGACCATCATCTTTAAAGTGAACAACCTTTA	Majority
	9810 9820 9830 9840 9850	
9789	TAGATTTTAAACTTCTGTTTGACCATCATCTTTAAAGTGAACAACCTTTA	cohl_a12.seq
9786	TAGATTTTAAACTTCTGTTTGACCATCATCTTTAAAGTGAACAACCTTTA	a909_a12.seq
	AGGTTTTTCATCTGAAGCTTCTAATGGCTTATCATAAGTTGACCTCTACTTT	Majority
	9860 9870 9880 9890 9900	
9839	AGGTTTTTCATCTGAAGCTTCTAATGGCTTATCATAAGTTGACCTCTACTTT	cohl_a12.seq
9836	AGGTTTTTCATCTGAAGCTTCTAATGGCTTATCATAAGTTGACCTCTACTTT	a909_a12.seq
	TACTGGGGCTTGGGGTTCTGCTTCTTTACCATTTGACTCAATAGTAATGT	Majority
	9910 9920 9930 9940 9950	
9889	TACTGGGGCTTGGGGTTCTGCTTCTTTACCATTTGACTCAATAGTAATGT	cohl_a12.seq
9886	TACTGGGGCTTGGGGTTCTGCTTCTTTACCATTTGACTCAATAGTAATGT	a909_a12.seq
	CATAGAGTTTGAAGTTTTTGATTTCACTATCTTGTTTAGCAACTTCTGTC	Majority
	9960 9970 9980 9990 10000	
9939	CATAGAGTTTGAAGTTTTTGATTTCACTATCTTGTTTAGCAACTTCTGTC	cohl_a12.seq
9936	CATAGAGTTTGAAGTTTTTGATTTCACTATCTTGTTTAGCAACTTCTGTC	a909_a12.seq
	AATGCTTTTTTCTTATAGTCTTTTAAAGTAGCTGAATTGCTTTTTAATTC	Majority
	10010 10020 10030 10040 10050	
9989	AATGCTTTTTTCTTATAGTCTTTTAAAGTAGCTGAATTGCTTTTTAATTC	cohl_a12.seq
9986	AATGCTTTTTTCTTATAGTCTTTTAAAGTAGCTGAATTGCTTTTTAATTC	a909_a12.seq
	CGTCACCTTTAAATCAGCATTTTATAGGAATCTTAGCTTCTTTGGTCAAAG	Majority
	10060 10070 10080 10090 10100	
10039	CGTCACCTTTAAATCAGCATTTTATAGGAATCTTAGCTTCTTTGGTCAAAG	cohl_a12.seq
10036	CGTCACCTTTAAATCAGCATTTTATAGGAATCTTAGCTTCTTTGGTCAAAG	a909_a12.seq
	TCACTGTTACAGTATAGTCTGCACCTCTAAACATCAATGGTTCTTTCACGG	Majority
	10110 10120 10130 10140 10150	
10089	TCACTGTTACAGTATAGTCTGCACCTCTAAACATCAATGGTTCTTTCACGG	cohl_a12.seq
10086	TCACTGTTACAGTATAGTCTGCACCTCTAAACATCAATGGTTCTTTCACGG	a909_a12.seq
	TAAGCAGCTTCTCAGAAAGATGATGTTTCTGTTACACTAGAAAGCAGGAGT	Majority
	10160 10170 10180 10190 10200	
10139	TAAGCAGCTTCTCAGAAAGATGATGTTTCTGTTACACTAGAAAGCAGGAGT	cohl_a12.seq
10136	TAAGCAGCTTCTCAGAAAGATGATGTTTCTGTTACACTAGAAAGCAGGAGT	a909_a12.seq
	CTGTGGCTTGGCTCTGCTCAACACTTGATTGAGAACTAGATGTTGATGAAG	Majority
	10210 10220 10230 10240 10250	
10189	CTGTGGCTTGGCTCTGCTCAACACTTGATTGAGAACTAGATGTTGATGAAG	cohl_a12.seq
10186	CTGTGGCTTGGCTCTGCTCAACACTTGATTGAGAACTAGATGTTGATGAAG	a909_a12.seq
	TTACCTGGCTAGAAATTTTATTTTCTAAAGTAATCCCAACATCATCTGTC	Majority
	10260 10270 10280 10290 10300	
10239	TTACCTGGCTAGAAATTTTATTTTCTAAAGTAATCCCAACATCATCTGTC	cohl_a12.seq
10236	TTACCTGGCTAGAAATTTTATTTTCTAAAGTAATCCCAACATCATCTGTC	a909_a12.seq
	TTAGTTTCTTCAACTGTTATTGCTGCTAGAAATTAATAAATAAGTCGTTAA	Majority
	10310 10320 10330 10340 10350	
10289	TTAGTTTCTTCAACTGTTATTGCTGCTAGAAATTAATAAATAAGTCGTTAA	cohl_a12.seq
10286	TTAGTTTCTTCAACTGTTATTGCTGCTAGAAATTAATAAATAAGTCGTTAA	a909_a12.seq
	AAAAGTTGTTAGGATCATCAATGACCACATGATAATTTTCCACTCTTTAG	Majority
	10360 10370 10380 10390 10400	
10339	AAAAGTTGTTAGGATCATCAATGACCACATGATAATTTTCCACTCTTTAG	cohl_a12.seq
10336	AAAAGTTGTTAGGATCATCAATGACCACATGATAATTTTCCACTCTTTAG	a909_a12.seq

FIGURE 210

**FIGURE 21P**

127/487

CGCCTCTCTTATCTACTAAATTTCTTTACGTTTTCTAAATAGGTATCTGCA Majority										
	11060	11070	11080	11090	11100					
11039	CGCCTCTCTTATCTACTAAATTTCTTTACGTTTTCTAAATAGGTATCTGCA coh1_a12.seq									
11035	CGCCTCTCTTATCTACTAAATTTCTTTACGTTTTCTAAATAGGTATCTGCA a909_a12.seq									
GCTAACTGAGCAGTCTCAAAAAATACCACTAAGCATTAAAGGAAGCTTCGGC Majority										
	11110	11120	11130	11140	11150					
11089	GCTAACTGAGCAGTCTCAAAAAATACCACTAAGCATTAAAGGAAGCTTCGGC coh1_a12.seq									
11085	GCTAACTGAGCAGTCTCAAAAAATACCACTAAGCATTAAAGGAAGCTTCGGC a909_a12.seq									
AATAGAACCAGCTTTGCTAATTTTGAATTTTTTATCATCTAGCGCTTCTT Majority										
	11160	11170	11180	11190	11200					
11139	AATAGAACCAGCTTTGCTAATTTTGAATTTTTTATCATCTAGCGCTTCTT coh1_a12.seq									
11135	AATAGAACCAGCTTTGCTAATTTTGAATTTTTTATCATCTAGCGCTTCTT a909_a12.seq									
TAAGTTGCTGAATCTCTTTCTCTTGTTTTTCAATAAGAAGTTGCTGTTCT Majority										
	11210	11220	11230	11240	11250					
11189	TAAGTTGCTGAATCTCTTTCTCTTGTTTTTCAATAAGAAGTTGCTGTTCT coh1_a12.seq									
11185	TAAGTTGCTGAATCTCTTTCTCTTGTTTTTCAATAAGAAGTTGCTGTTCT a909_a12.seq									
AACATAAATTTCTAGCAAGTCTTTTCTTTTAAATTTTTTAAATCTTCCAT Majority										
	11260	11270	11280	11290	11300					
11239	AACATAAATTTCTAGCAAGTCTTTTCTTTTAAATTTTTTAAATCTTCCAT coh1_a12.seq									
11235	AACATAAATTTCTAGCAAGTCTTTTCTTTTAAATTTTTTAAATCTTCCAT a909_a12.seq									
CGCGATTACTTCTTAACCTTAAATTATCGTTTAGATATTATATC Majority										
	11310	11320	11330	11340	11350					
11289	CGCGATTACTTCTTAACCTTAAATTATCGTTTAGATATTATATC coh1_a12.seq									
11285	CGCGATTACTTCTTAACCTTAAATTATCGTTTAGATATTATATC a909_a12.seq									
AAAGTTCTAACCTTTAAACTCATTTTTTGTCTGTTTTTTCTCAAAAAA Majority										
	11360	11370	11380	11390	11400					
11339	AAAGTTCTAACCTTTAAACTCATTTTTTGTCTGTTTTTTCTCAAAAAA coh1_a12.seq									
11335	AAAGTTCTAACCTTTAAACTCATTTTTTGTCTGTTTTTTCTCAAAAAA a909_a12.seq									
AGTCTATGCTAAATTAACATTTTTGATAATTTTTTGAAAAATCTCATCGA Majority										
	11410	11420	11430	11440	11450					
11389	AGTCTATGCTAAATTAACATTTTTGATAATTTTTTGAAAAATCTCATCGA coh1_a12.seq									
11385	AGTCTATGCTAAATTAACATTTTTGATAATTTTTTGAAAAATCTCATCGA a909_a12.seq									
AGTCATTTTTCTTTTGAAGCTCGAATTCTAGGCATTAAAAAGCCATATA Majority										
	11460	11470	11480	11490	11500					
11439	AGTCATTTTTCTTTTGAAGCTCGAATTCTAGGCATTAAAAAGCCATATA coh1_a12.seq									
11435	AGTCATTTTTCTTTTGAAGCTCGAATTCTAGGCATTAAAAAGCCATATA a909_a12.seq									
TCAAAATTGATATATGGCTTTTTTTATTATTTAAACAAAAGCAATCAATAG Majority										
	11510	11520	11530	11540	11550					
11489	TCAAAATTGATATATGGCTTTTTTTATTATTTAAACAAAAGCAATCAATAG coh1_a12.seq									
11476	TCAAAATTGATATATGGCTTTTTTTATTATTTAAACAAAAGCAATCAATAG a909_a12.seq									
GACAATAGCGTCAATTTAGTGACATAATCTATTACAGATTAAAGTTCTTTT Majority										
	11560	11570	11580	11590	11600					
11539	GACAATAGCGTCAATTTAGTGACATAATCTATTACAGATTAAAGTTCTTTT coh1_a12.seq									
11492	GACAATAGCGTCAATTTAGTGACATAATCTATTACAGATTAAAGTTCTTTT a909_a12.seq									
TGAATAATATAATCCAACCTTTTCAACTGTTTTTCCCATGTGAAATGTTT Majority										
	11610	11620	11630	11640	11650					
11589	TGAATAATATAATCCAACCTTTTCAACTGTTTTTCCCATGTGAAATGTTT coh1_a12.seq									
11542	TGAATAATATAATCCAACCTTTTCAACTGTTTTTCCCATGTGAAATGTTT a909_a12.seq									
TTTAATTCTTTTAGCAATATTCTGTTGTAGTTTCTCTCTTAATGCCTTAT Majority										
	11660	11670	11680	11690	11700					
11639	TTTAATTCTTTTAGCAATATTCTGTTGTAGTTTCTCTCTCTTAATGCCTTAT coh1_a12.seq									
11592	TTTAATTCTTTTAGCAATATTCTGTTGTAGTTTCTCTCTCTTAATGCCTTAT a909_a12.seq									

FIGURE 21Q

128/487

```

CTTTTAACTAAATAAATCAAGAGATTTCATGAGTGAAGTATTTTCTTCC Majority
11710 11720 11730 11740 11750
11689 CTTTACTAAATAAATCAAGAGATTTCATGAGTGAAGTATTTTCTTCC cohl_a12.seq
11642 CTTTACTAAATAAATCAAGAGATTTCATGAGTGAAGTATTTTCTTCC a909_a12.seq

ATGATGATTCTTAACCTCAGGGCTATCAATAAATTCAACTGTTCCACCGCC Majority
11760 11770 11780 11790 11800
11739 ATGATGATTCTTAACCTCAGGGCTATCAATAAATTCAACTGTTCCACCGCC cohl_a12.seq
11692 ATGATGATTCTTAACCTCAGGGCTATCAATAAATTCAACTGTTCCACCGCC a909_a12.seq

ATCTGTTGCAATAATAGCACTCGAAAGTAGACCAGCTTCTAAAAATAGAGG Majority
11810 11820 11830 11840 11850
11789 ATCTGTTGCAATAATAGCACTCGAAAGTAGACCAGCTTCTAAAAATAGAGG cohl_a12.seq
11742 ATCTGTTGCAATAATAGCACTCGAAAGTAGACCAGCTTCTAAAAATAGAGG a909_a12.seq

TTGGTAATCCCTCTCGGATACATTGATGGGTAAACAAAGATATCTGTCTGT Majority
11860 11870 11880 11890 11900
11839 TTGGTAATCCCTCTCGGATACATTGATGGGTAAACAAAGATATCTGTCTGT cohl_a12.seq
11792 TTGGTAATCCCTCTCGGATACATTGATGGGTAAACAAAGATATCTGTCTGT a909_a12.seq

GCCATTAAAGACATAGTCTGTTCAAAGTTTAAATTTCCCAAAAAAGTTAAT Majority
11910 11920 11930 11940 11950
11889 GCCATTAAAGACATAGTCTGTTCAAAGTTTAAATTTCCCAAAAAAGTTAAT cohl_a12.seq
11842 GCCATTAAAGACATAGTCTGTTCAAAGTTTAAATTTCCCAAAAAAGTTAAT a909_a12.seq

CTGTTTGCACTGATATTTCTCTTTCAAATGTGCTAATTCAGGTCCGTCTC Majority
11960 11970 11980 11990 12000
11939 CTGTTTGCACTGATATTTCTCTTTCAAATGTGCTAATTCAGGTCCGTCTC cohl_a12.seq
11892 CTGTTTGCACTGATATTTCTCTTTCAAATGTGCTAATTCAGGTCCGTCTC a909_a12.seq

CTGCAATCTGTAAATAAACATTTTTCAGAGTACTGTGACATCGAAAAATGCT Majority
12010 12020 12030 12040 12050
11989 CTGCAATCTGTAAATAAACATTTTTCAGAGTACTGTGACATCGAAAAATGCT cohl_a12.seq
11942 CTGCAATCTGTAAATAAACATTTTTCAGAGTACTGTGACATCGAAAAATGCT a909_a12.seq

TCTAAGAGCAATTCAATGCCCTTTTCTTTAATAATTCTACCAGCATAAGT Majority
12060 12070 12080 12090 12100
12039 TCTAAGAGCAATTCAATGCCCTTTTCTTTAATAATTCTACCAGCATAAGT cohl_a12.seq
11992 TCTAAGAGCAATTCAATGCCCTTTTCTTTAATAATTCTACCAGCATAAGT a909_a12.seq

GATGAAAAATATCATCAGCAGATTTTTCAGGTAAGCCGTGTCAGCAAAAT Majority
12110 12120 12130 12140 12150
12089 GATGAAAAATATCATCAGCAGATTTTTCAGGTAAGCCGTGTCAGCAAAAT cohl_a12.seq
12042 GATGAAAAATATCATCAGCAGATTTTTCAGGTAAGCCGTGTCAGCAAAAT a909_a12.seq

CAGAGCCTAGACTTTTCAGATACCGAATTATAAAATAACTCTTTAGCTTCT Majority
12160 12170 12180 12190 12200
12139 CAGAGCCTAGACTTTTCAGATACCGAATTATAAAATAACTCTTTAGCTTCT cohl_a12.seq
12092 CAGAGCCTAGACTTTTCAGATACCGAATTATAAAATAACTCTTTAGCTTCT a909_a12.seq

ATATTTAAATGTTTTAACCATTCAACGGCTTCTCTTGGATACCGCATAAAA Majority
12210 12220 12230 12240 12250
12189 ATATTTAAATGTTTTAACCATTCAACGGCTTCTCTTGGATACCGCATAAAA cohl_a12.seq
12142 ATATTTAAATGTTTTAACCATTCAACGGCTTCTCTTGGATACCGCATAAAA a909_a12.seq

ATCTGGACCATAGTCTTAACACGGCGCTGTGAGAGATGTTTCATAGATAG Majority
12260 12270 12280 12290 12300
12239 ATCTGGACCATAGTCTTAACACGGCGCTGTGAGAGATGTTTCATAGATAG cohl_a12.seq
12192 ATCTGGACCATAGTCTTAACACGGCGCTGTGAGAGATGTTTCATAGATAG a909_a12.seq

CTCCAAAGAAATCTAAAAAACGATTATTGACAGAAAAATGACTTGACCCA Majority
12310 12320 12330 12340 12350
12289 CTCCAAAGAAATCTAAAAAACGATTATTGACAGAAAAATGACTTGACCCA cohl_a12.seq
12242 CTCCAAAGAAATCTAAAAAACGATTATTGACAGAAAAATGACTTGACCCA a909_a12.seq

```

FIGURE 21R

129/487

	TGGTCTAAAAACAATACTAGGTAACGGTAAACGGTGGCTTCTTTGCAAAACATAGCCC	Majority
	12360 12370 12380 12390 12400	
12339	TGGTCTAAAAACAATACTAGGTAACGGTGGCTTCTTTGCAAAACATAGCCC	cohl_a12.seq
12292	TGGTCTAAAAACAATACTAGGTAACGGTGGCTTCTTTGCAAAACATAGCCC	a909_a12.seq
	TTCTAGCGTTGTTAACTGAAAACGGTGTATTACAAATCACAAAATCAATAT	Majority
	12410 12420 12430 12440 12450	
12389	TTCTAGCGTTGTTAACTGAAAACGGTGTATTACAAATCACAAAATCAATAT	cohl_a12.seq
12342	TTCTAGCGTTGTTAACTGAAAACGGTGTATTACAAATCACAAAATCAATAT	a909_a12.seq
	TTTCATCTGAAAACATGTTTCATCAGCGTGTGTTGTTCTCGATTTTTTGTTA	Majority
	12460 12470 12480 12490 12500	
12439	TTTCATCTGAAAACATGTTTCATCAGCGTGTGTTGTTCTCGATTTTTTGTTA	cohl_a12.seq
12392	TTTCATCTGAAAACATGTTTCATCAGCGTGTGTTGTTCTCGATTTTTTGTTA	a909_a12.seq
	ATAATAGGATAGCGCTGCTTGACAAAGTTTGGTTCGGTAAACCGGTAAAT	Majority
	12510 12520 12530 12540 12550	
12489	ATAATAGGATAGCGCTGCTTGACAAAGTTTGGTTCGGTAAACCGGTAAAT	cohl_a12.seq
12442	ATAATAGGATAGCGCTGCTTGACAAAGTTTGGTTCGGTAAACCGGTAAAT	a909_a12.seq
	TTTTCTACCCCTTGCTTTCATCTATAATCGGTAAATCATCATGATTAGTTG	Majority
	12560 12570 12580 12590 12600	
12539	TTTTCTACCCCTTGCTTTCATCTATAATCGGTAAATCATCATGATTAGTTG	cohl_a12.seq
12492	TTTTCTACCCCTTGCTTTCATCTATAATCGGTAAATCATCATGATTAGTTG	a909_a12.seq
	TTACAATAACAACACGGTAGCCACGGCTTAACCAATCTGCTGTCAATTTTA	Majority
	12610 12620 12630 12640 12650	
12589	TTACAATAACAACACGGTAGCCACGGCTTAACCAATCTGCTGTCAATTTTA	cohl_a12.seq
12542	TTACAATAACAACACGGTAGCCACGGCTTAACCAATCTGCTGTCAATTTTA	a909_a12.seq
	TCTGTATAACGTTCAATACCTCCGAGGAAGGGTAGATAATATCCTGAGAA	Majority
	12660 12670 12680 12690 12700	
12639	TCTGTATAACGTTCAATACCTCCGAGGAAGGGTAGATAATATCCTGAGAA	cohl_a12.seq
12592	TCTGTATAACGTTCAATACCTCCGAGGAAGGGTAGATAATATCCTGAGAA	a909_a12.seq
	AACAGCAACTGTTTTTACCTTATTTTCCATATTTTATCCACTTTTCATCAAT	Majority
	12710 12720 12730 12740 12750	
12689	AACAGCAACTGTTTTTACCTTATTTTCCATATTTTATCCACTTTTCATCAAT	cohl_a12.seq
12642	AACAGCAACTGTTTTTACCTTATTTTCCATATTTTATCCACTTTTCATCAAT	a909_a12.seq
	AAGCCATCTTTTTAAGCCTTTAATCATAGCAACTAATTTTTTGGCTCTTTTG	Majority
	12760 12770 12780 12790 12800	
12739	AAGCCATCTTTTTAAGCCTTTAATCATAGCAACTAATTTTTTGGCTCTTTTG	cohl_a12.seq
12692	AAGCCATCTTTTTAAGCCTTTAATCATAGCAACTAATTTTTTGGCTCTTTTG	a909_a12.seq
	CTCTTCTGCTACCAACACTCGAACAATTCATTTEGGCATAAAATACTAAAT	Majority
	12810 12820 12830 12840 12850	
12789	CTCTTCTGCTACCAACACTCGAACAATTCATTTEGGCATAAAATACTAAAT	cohl_a12.seq
12742	CTCTTCTGCTACCAACACTCGAACAATTCATTTEGGCATAAAATACTAAAT	a909_a12.seq
	ATTGTGCGCGCTTCTTCTTACCATATTTTATTAATATAAATCGCATTC	Majority
	12860 12870 12880 12890 12900	
12839	ATTGTGCGCGCTTCTTCTTACCATATTTTATTAATATAAATCGCATTC	cohl_a12.seq
12792	ATTGTGCGCGCTTCTTCTTACCATATTTTATTAATATAAATCGCATTC	a909_a12.seq
	CGTATCATGTAATATTTTTCGAAATGGTGAATGATTCAATACATGAAAAAC	Majority
	12910 12920 12930 12940 12950	
12889	CGTATCATGTAATATTTTTCGAAATGGTGAATGATTCAATACATGAAAAAC	cohl_a12.seq
12842	CGTATCATGTAATATTTTTCGAAATGGTGAATGATTCAATACATGAAAAAC	a909_a12.seq
	ATGCCCAAATTTTTTAACTCGTGAAGAGTGTCCAATTTTCGTGTAAAAGAC	Majority
	12960 12970 12980 12990 13000	
12939	ATGCCCAAATTTTTTAACTCGTGAAGAGTGTCCAATTTTCGTGTAAAAGAC	cohl_a12.seq
12892	ATGCCCAAATTTTTTAACTCGTGAAGAGTGTCCAATTTTCGTGTAAAAGAC	a909_a12.seq

FIGURE 21S



130/487

	C A A T A T A A T T A A C C T G A T A A G T C T T A T A T C C C A T C T C T G A C A G A C G A T A A	Majority
	13010 13020 13030 13040 13050	
12989	C A A T A A A A T T A A C C T G A T A A G T C T T A T A T C C C A T C T C T G A C A G A C G A T A A	cohl_al2.seq
12942	C A A T A A A A T T A A C C T G A T A A G T C T T A T A T C C C A T C T C T G A C A G A C G A T A A	a909_al2.seq
	T T C A T T T C A G A G T C A A C A A A A T C A A T A A A C A T C T C T T C T G C A A A G C C A G A	Majority
	13060 13070 13080 13090 13100	
13039	T T C A T T T C A G A G T C A A C A A A A T C A A T A A A C A T C T C T T C T G C A A A G C C A G A	cohl_al2.seq
12992	T T C A T T T C A G A G T C A A C A A A A T C A A T A A A C A T C T C T T C T G C A A A G C C A G A	a909_al2.seq
	T G T T T C T T C A A A A A C G C T C G T T T T C A T T A A A G C A G C C G A A G T A A T A C A C T	Majority
	13110 13120 13130 13140 13150	
13089	T G T T T C T T C A A A A A C G C T C G T T T T C A T T A A A G C A G C C G A A G T A A T A C A C T	cohl_al2.seq
13042	T G T T T C T T C A A A A A C G C T C G T T T T C A T T A A A G C A G C C G A A G T A A T A C A C T	a909_al2.seq
	C T T C A A T T T C T T T A T A G T C A A A T T C T T G C A T C A C T A A A T C T T C A C G G T T C	Majority
	13160 13170 13180 13190 13200	
13139	C T T C A A T T T C T T T A T A G T C A A A T T C T T G C A T A C T A A A T C T T C A C G G T T C	cohl_al2.seq
13092	C T T C A A T T T C T T T A T A G T C A A A T T C T T G C A T C A C T A A A T C T T C A C G G T T C	a909_al2.seq
	A T A T C T T G A T A C A A A C A A G A T A A C A T A C C G A C C T T A G G T A A A T G A A G G T A	Majority
	13210 13220 13230 13240 13250	
13189	A T A T C T T G A T A C A A A C A A G A T A A C A T A C C G A C C T T A G G T A A A T G A A G G T A	cohl_al2.seq
13142	A T A T C T T G A T A C A A A C A A G A T A A C A T A C C G A C C T T A G G T A A A T G A A G G T A	a909_al2.seq
	A T T T T C A T A A T T A T C T A T C A A A T C A C C T A G C A C A C C G A A T C T T G A T C T A	Majority
	13260 13270 13280 13290 13300	
13239	A T T T T C A T A A T T A T C T A T C A A A T C A C C T A G C A C A C C G A A T C T T G A T C T A	cohl_al2.seq
13192	A T T T T C A T A A T T A T C T A T C A A A T C A C C T A G C A C A C C G A A T C T T G A T C T A	a909_al2.seq
	A A G T C A A G A A C C A A T C A A A A T T C T T G T G C T A C T G C A A A T T G A C C G A T A C A G	Majority
	13310 13320 13330 13340 13350	
13289	A A G T C A A G A A C C A A T C A A A A T T C T T G T G C T A C T G C A A A T T G A C C G A T A C A G	cohl_al2.seq
13242	A A G T C A A G A A C C A A T C A A A A T T C T T G T G C T A C T G C A A A T T G A C C A T A C A G	a909_al2.seq
	T T C A A A G C A T A T G C A A T T C C T T T A T T T T C T G T T A A A T A A T C A A C A G T T A G	Majority
	13360 13370 13380 13390 13400	
13339	T T C A A A G C A T A T G C A A T T C C T T T A T T T T C T G T T A A A T A A T C A A C A G T T A G	cohl_al2.seq
13292	T T C A A A G C A T A T G C A A T T C C T T T A T T T T C T G T T A A A T A A T C A A C A G T T A G	a909_al2.seq
	G T G C C C C T C T T C A T T A T A A T C G G C C A C T A A T T G A G A A A T T T C T T C C T T A T	Majority
	13410 13420 13430 13440 13450	
13389	G T G C C C C T C T T C A T T A T A A T C G G C C A C T A A T T G A G A A A T T T C T T C C T T A T	cohl_al2.seq
13342	G T G C C C C T C T T C A T T A T A A T C G G C C A C T A A T T G A G A A A T T T C T T C C T T A T	a909_al2.seq
	T T T T C G A G C C A T T A T C T A C C A T G T A G A T A T G G C T T A C T T G A G G A T A A A T T	Majority
	13460 13470 13480 13490 13500	
13439	T T T T C G A G C C A T T A T C T A C C A T G T A G A T A T G G C T T A C T T G A G G A T A A A T T	cohl_al2.seq
13392	T T T T C G A G C C A T T A T C T A C C A T G T A G A T A T G G C T T A C T T G A G G A T A A A T T	a909_al2.seq
	G C T C G A A T G T T C T G A T G T A A G C G T T C A A T A T T G C G G T T A A A G G T G A C A A T	Majority
	13510 13520 13530 13540 13550	
13489	G C T C G A A T G T T C T G A T G T A A G C G T T C A A T A T T G G G G T T A A A G G T G A C A A T	cohl_al2.seq
13442	G C T C G A A T G T T C T G A T G T A A G C G T T C A A T A T T G G G G T T A A A G G T G A C A A T	a909_al2.seq
	A C C C G C T A A A T A T T T C A T G T T C T A T G C T C T T T T C T A A A A T C T C T A A A T A A	Majority
	13560 13570 13580 13590 13600	
13539	A C C C G C T A A A T A T T T C A T G T T C T A T G C T C T T T T C T A A A A T C T C T A A A T A A	cohl_al2.seq
13492	A C C C G C T A A A T A T T T C A T G T T C T A T G C T C T T T T C T A A A A T C T C T A A A T A A	a909_al2.seq
	C T G A A T G A C T G G T G C T T T G G T T A T A A A A C G A T A C C G A C A T A G A T A G T T A	Majority
	13610 13620 13630 13640 13650	
13589	C T G A A T G A C T G G T G C T T T G G T T A T A A A A C G A T A C C G A C A T A G A T A G T T A	cohl_al2.seq
13542	C T G A A T G A C T G G T G C T T T G G T T A T A A A A C G A T A C C G A C A T A G A T A G T T A	a909_al2.seq

FIGURE 21T

131/487

	CTGCT	13660	13670	13680	13690	13700	Majority
13639	CTGCTACTAAACTTTTGAATGACATAATTTACCAATGATACTGACATTTTGA						coh1_al2.seq
13592	CTGCTACTAAACTTTTGAATGACATAATTTACCAATGATACTGACATTTTGA						a909_al2.seq
	GTATTGATATAATAGAGTACAGCTCCACTAAGAGTAGCAGCAATTAAATA						Majority
		13710	13720	13730	13740	13750	
13689	GTATTGATATAATAAGTACAGCTCCACTAAGAGTAGCAGCAATTAAATA						coh1_al2.seq
13642	GTATTGATATAATAGAGTACAGCTCCACTAAGAGTAGCAGCAATTAAATA						a909_al2.seq
	GCCGAGCATTTCCTCTTTGTTAATTCTTTTAAAGTAAATACATCTCTTAAAG						Majority
		13760	13770	13780	13790	13800	
13739	GCCGAGCATTTCCTCTTTGTTAATTCTTTTAAAGTAAATACATCTCTTAAAG						coh1_al2.seq
13692	GCCGAGCATTTCCTCTTTGTTAATTCTTTTAAAGTAAATACATCTCTTAAAG						a909_al2.seq
	AGATAGCTTTCATATAGGGAGACAATAAATTCAGTAATAACTGTAGAGATA						Majority
		13810	13820	13830	13840	13850	
13789	AGATAGCTTTCATATAGGGAGACAATAAATTCAGTAATAACTGTAGAGATA						coh1_al2.seq
13742	AGATAGCTTTCATATAGGGAGACAATAAATTCAGTAATAACTGTAGAGATA						a909_al2.seq
	ATAGCTCCCATAGCACCTAAATTTGGTATTAAAAAGTATATTAAGCACAAAC						Majority
		13860	13870	13880	13890	13900	
13839	ATAGCTCCCATAGCACCTAAATTTGGTATTAAAAAGTATATTAAGCACAAAC						coh1_al2.seq
13792	ATAGCTCCCATAGCACCTAAATTTGGTATTAAAAAGTATATTAAGCACAAAC						a909_al2.seq
	ATTTGCCACAAGTCCAATAACTGCAGACATTGTGTAAGCTTTTGTACGTC						Majority
		13910	13920	13930	13940	13950	
13889	ATTTGCCACAAGTCCAATAACTGCAGACATTGTGTAAGCTTTTGTACGTC						coh1_al2.seq
13842	ATTTGCCACAAGTCCAATAACTGCAGACATTGTGTAAGCTTTTGTACGTC						a909_al2.seq
	TTGAAGCCAGTAGATACTGTGTCCCTAAAGCGTTACCATAAGAAATGCAA						Majority
		13960	13970	13980	13990	14000	
13939	TTGAAGCCAGTAGATACTGTGTCCCTAAAGCGTTACCATAAGAAATGCAA						coh1_al2.seq
13892	TTGAAGCCAGTAGATACTGTGTCCCTAAAGCGTTACCATAAGAAATGCAA						a909_al2.seq
	ATGATCATCAAA						Majority
		14010					
13989	ATGATCATCAAA						coh1_al2.seq
13942	ATGATCATCAAA						a909_al2.seq

Decoration 'Decoration #1': Shade (with solid black) residues that differ from the Consensus.

FIGURE 21U

132/487

M K L S K K L L F S A A V L T M V A G S T V E P V A Q F A T G M S I V R A A E V S Q E R P A K T T V																																	Majority																			
10											20											30											40										50									
1	M	K	L	S	K	K	L	L	F	S	A	A	V	L	T	M	V	A	G	S	T	V	E	P	V	A	Q	F	A	T	G	M	S	I	V	R	A	A	E	V	S	Q	E	R	P	A	K	T	T	V	sag645_2603.pep	
1	M	K	L	S	K	K	L	L	F	S	A	A	V	L	T	M	V	A	G	S	T	V	E	P	V	A	Q	F	A	T	G	M	S	I	V	R	A	A	E	V	S	Q	E	R	P	A	K	T	T	V	sag645_a909.pep	
1	M	K	L	S	K	K	L	L	F	S	A	A	V	L	T	M	V	A	G	S	T	V	E	P	V	A	Q	F	A	T	G	M	S	I	V	R	A	A	E	V	S	Q	E	R	P	A	K	T	T	V	sag645_cjb111.pep	
1	M	K	L	S	K	K	L	L	F	S	A	A	V	L	T	M	V	A	G	S	T	V	E	P	V	A	Q	F	A	T	G	M	S	I	V	R	A	A	E	V	S	Q	E	R	P	A	K	T	T	V	sag645_cohl.pep	
1	M	K	L	S	K	K	L	L	F	S	A	A	V	L	T	M	V	A	G	S	T	V	E	P	V	A	Q	F	A	T	G	M	S	I	V	R	A	A	E	V	S	Q	E	R	P	A	K	T	T	V	sag645_nem316.pep	
N I Y K L Q A D S Y K S E I T S N G G I E N K D G E V I S N Y A K L G D N V K G L Q C V Q F K R Y K																																	Majority																			
60											70											80											90										100									
51	N	I	Y	K	L	Q	A	D	S	Y	K	S	E	I	T	S	N	G	G	I	E	N	K	D	G	E	V	I	S	N	Y	A	K	L	G	D	N	V	K	G	L	Q	C	V	Q	F	K	R	Y	K	sag645_2603.pep	
51	N	I	Y	K	L	Q	A	D	S	Y	K	S	E	I	T	S	N	G	G	I	E	N	K	D	G	E	V	I	S	N	Y	A	K	L	G	D	N	V	K	G	L	Q	C	V	Q	F	K	R	Y	K	sag645_a909.pep	
51	N	I	Y	K	L	Q	A	D	S	Y	K	S	E	I	T	S	N	G	G	I	E	N	K	D	G	E	V	I	S	N	Y	A	K	L	G	D	N	V	K	G	L	Q	C	V	Q	F	K	R	Y	K	sag645_cjb111.pep	
51	N	I	Y	K	L	Q	A	D	S	Y	K	S	E	I	T	S	N	G	G	I	E	N	K	D	G	E	V	I	S	N	Y	A	K	L	G	D	N	V	K	G	L	Q	C	V	Q	F	K	R	Y	K	sag645_cohl.pep	
51	N	I	Y	K	L	Q	A	D	S	Y	K	S	E	I	T	S	N	G	G	I	E	N	K	D	G	E	V	I	S	N	Y	A	K	L	G	D	N	V	K	G	L	Q	C	V	Q	F	K	R	Y	K	sag645_nem316.pep	
V K T D I S V D E L K K L T T V E A A D A K V G T I L E E G V S L P Q K T N A Q C L V D A L D S K																																	Majority																			
110											120											130											140										150									
01	V	K	T	D	I	S	V	D	E	L	K	K	L	T	T	V	E	A	A	D	A	K	V	G	T	I	L	E	E	G	V	S	L	P	Q	K	T	N	A	Q	C	L	V	D	A	L	D	S	K	sag645_2603.pep		
01	V	K	T	D	I	S	V	D	E	L	K	K	L	T	T	V	E	A	A	D	A	K	V	G	T	I	L	E	E	G	V	S	L	P	Q	K	T	N	A	Q	C	L	V	D	A	L	D	S	K	sag645_a909.pep		
01	V	K	T	D	I	S	V	D	E	L	K	K	L	T	T	V	E	A	A	D	A	K	V	G	T	I	L	E	E	G	V	S	L	P	Q	K	T	N	A	Q	C	L	V	D	A	L	D	S	K	sag645_cjb111.pep		
01	V	K	T	D	I	S	V	D	E	L	K	K	L	T	T	V	E	A	A	D	A	K	V	G	T	I	L	E	E	G	V	S	L	P	Q	K	T	N	A	Q	C	L	V	D	A	L	D	S	K	sag645_nem316.pep		
S N V R Y L Y V E D L K N S P S N I T K A Y A V P F V L E L P V A N S T G T G F L S E I N I Y P K N																																	Majority																			
160											170											180											190										200									
51	S	N	V	R	Y	L	Y	V	E	D	L	K	N	S	P	S	N	I	T	K	A	Y	A	V	P	F	V	L	E	L	P	V	A	N	S	T	G	T	G	F	L	S	E	I	N	I	Y	P	K	N	sag645_2603.pep	
51	S	N	V	R	Y	L	Y	V	E	D	L	K	N	S	P	S	N	I	T	K	A	Y	A	V	P	F	V	L	E	L	P	V	A	N	S	T	G	T	G	F	L	S	E	I	N	I	Y	P	K	N	sag645_a909.pep	
51	S	N	V	R	Y	L	Y	V	E	D	L	K	N	S	P	S	N	I	T	K	A	Y	A	V	P	F	V	L	E	L	P	V	A	N	S	T	G	T	G	F	L	S	E	I	N	I	Y	P	K	N	sag645_cjb111.pep	
51	S	N	V	R	Y	L	Y	V	E	D	L	K	N	S	P	S	N	I	T	K	A	Y	A	V	P	F	V	L	E	L	P	V	A	N	S	T	G	T	G	F	L	S	E	I	N	I	Y	P	K	N	sag645_cohl.pep	
51	S	N	V	R	Y	L	Y	V	E	D	L	K	N	S	P	S	N	I	T	K	A	Y	A	V	P	F	V	L	E	L	P	V	A	N	S	T	G	T	G	F	L	S	E	I	N	I	Y	P	K	N	sag645_nem316.pep	
V V T D E P K T D K D V K K L G Q D D A G Y T I G E E F K W F L K S T I P A N L G D Y E K F E I T D																																	Majority																			
210											220											230											240										250									
01	V	V	T	D	E	P	K	T	D	K	D	V	K	K	L	G	Q	D	D	A	G	Y	T	I	G	E	E	F	K	W	F	L	K	S	T	I	P	A	N	L	G	D	Y	E	K	F	E	I	T	D	sag645_2603.pep	
01	V	V	T	D	E	P	K	T	D	K	D	V	K	K	L	G	Q	D	D	A	G	Y	T	I	G	E	E	F	K	W	F	L	K	S	T	I	P	A	N	L	G	D	Y	E	K	F	E	I	T	D	sag645_a909.pep	
01	V	V	T	D	E	P	K	T	D	K	D	V	K	K	L	G	Q	D	D	A	G	Y	T	I	G	E	E	F	K	W	F	L	K	S	T	I	P	A	N	L	G	D	Y	E	K	F	E	I	T	D	sag645_cjb111.pep	
01	V	V	T	D	E	P	K	T	D	K	D	V	K	K	L	G	Q	D	D	A	G	Y	T	I	G	E	E	F	K	W	F	L	K	S	T	I	P	A	N	L	G	D	Y	E	K	F	E	I	T	D	sag645_cohl.pep	
01	V	V	T	D	E	P	K	T	D	K	D	V	K	K	L	G	Q	D	D	A	G	Y	T	I	G	E	E	F	K	W	F	L	K	S	T	I	P	A	N	L	G	D	Y	E	K	F	E	I	T	D	sag645_nem316.pep	
K F A D G L T Y K S V G K I K I G S K T L N R D E H Y T I D E P T V D N Q N T L K I T F K P E K F K																																	Majority																			
260											270											280											290										300									
51	K	F	A	D	G	L	T	Y	K	S	V	G	K	I	K	I	G	S	K	T	L	N	R	D	E	H	Y	T	I	D	E	P	T	V	D	N	Q	N	T	L	K	I	T	F	K	P	E	K	F	K	sag645_2603.pep	
51	K	F	A	D	G	L	T	Y	K	S	V	G	K	I	K	I	G	S	K	T	L	N	R	D	E	H	Y	T	I	D	E	P	T	V	D	N	Q	N	T	L	K	I	T	F	K	P	E	K	F	K	sag645_a909.pep	
51	K	F	A	D	G	L	T	Y	K	S	V	G	K	I	K	I	G	S	K	T	L	N	R	D	E	H	Y	T	I	D	E	P	T	V	D	N	Q	N	T	L	K	I	T	F	K	P	E	K	F	K	sag645_cjb111.pep	
51	K	F	A	D	G	L	T	Y	K	S	V	G	K	I	K	I	G	S	K	T	L	N	R	D	E	H	Y	T	I	D	E	P	T	V	D	N	Q	N	T	L	K	I	T	F	K	P	E	K	F	K	sag645_cohl.pep	
51	K	F	A	D	G	L	T	Y	K	S	V	G	K	I	K	I	G	S	K	T	L	N	R	D	E	H	Y	T	I	D	E	P	T	V	D	N	Q	N	T	L	K	I	T	F	K	P	E	K	F	K	sag645_nem316.pep	
E I A E L L K G M T L V K N Q D A L D K A T A N T D D A A F L E I P V A S T I N E K A V L G K A I E																																	Majority																			
310											320											330											340										350									
11	E	I	A	E	L	L	K	G	M	T	L	V	K	N	Q	D	A	L	D	K	A	T	A	N	T	D	D	A	A	F	L	E	I	P	V	A	S	T	I	N	E	K	A	V	L	G	K	A	I	E	sag645_2603.pep	
11	E	I	A	E	L	L	K	G	M	T	L	V	K	N	Q	D	A	L	D	K	A	T	A	N	T	D	D	A	A	F	L	E	I	P	V	A	S	T	I	N	E	K	A	V	L	G	K	A	I	E	sag645_a909.pep	
11	E	I	A	E	L	L	K	G	M	T	L	V	K	N	Q	D	A	L	D	K	A	T	A	N	T	D	D	A	A	F	L	E	I	P	V	A	S	T	I	N	E	K	A	V	L	G	K	A	I	E	sag645_cjb111.pep	
11	E	I	A	E	L	L	K	G	M	T	L	V	K	N	Q	D	A	L	D	K	A	T	A	N	T	D	D	A	A	F	L	E	I	P	V	A	S	T	I	N	E	K	A	V	L	G	K	A	I	E	sag645_cohl.pep	
11	E	I	A	E	L	L	K	G	M	T	L	V	K	N	Q	D	A	L	D	K	A	T	A	N	T	D	D	A	A	F	L	E	I	P	V	A	S	T	I	N	E	K	A	V	L	G	K	A	I	E	sag645_nem316.pep	
N T F E L Q Y D H T P D K A D N P K P S N P P R K P E V H T G G K R F V K K D S T E T Q T L G G A E																																	Majority																			
360											370											380											390										400									
11	N	T	F	E	L	Q	Y	D	H	T	P	D	K	A	D	N	P	K	P	S	N	P	P	R	K	P	E	V	H	T	G	G	K	R	F	V	K	K	D	S	T	E	T	Q	T	L	G	G	A	E	sag645_2603.pep	
11	N	T	F	E	L	Q	Y	D	H	T	P	D	K	A	D	N	P	K	P	S	N	P	P	R	K	P	E	V	H	T	G	G	K	R	F	V	K	K	D	S	T	E	T	Q	T	L	G	G	A	E	sag645_a909.pep	
11	N	T	F	E	L	Q	Y	D	H	T	P	D	K	A	D	N	P	K	P	S	N	P	P	R	K	P	E	V	H	T	G	G	K	R	F	V	K	K	D	S	T	E	T	Q	T	L	G	G	A	E	sag645_cjb111.pep	
11	N	T	F	E	L	Q	Y	D	H	T	P	D	K	A	D	N	P	K	P	S	N	P	P	R	K	P	E	V	H	T	G	G	K	R	F	V	K	K	D	S	T	E	T	Q	T	L	G	G	A	E	sag645_cohl.pep	
11	N	T	F	E	L	Q	Y	D																																												

133/487

FDLLASDGTAVKWTDALEIKANTNKNYIACEAVTCQPIKLSHTDGTFEIK Majority									
410		420		430		440		450	
01	F	D	L	L	A	S	D	G	T
01	F	D	L	L	A	S	D	G	T
01	F	D	L	L	A	S	D	G	T
01	F	D	L	L	A	S	D	G	T
01	F	D	L	L	A	S	D	G	T
GLAYAVDANAEGTAVTYKLKETKAPECYVIPDKEIEFTVSQTSYNTKPTD Majority									
460		470		480		490		500	
51	G	L	A	Y	A	V	D	A	N
51	G	L	A	Y	A	V	D	A	N
51	G	L	A	Y	A	V	D	A	N
51	G	L	A	Y	A	V	D	A	N
51	G	L	A	Y	A	V	D	A	N
ITVDSADATPDTIKNNKRPSIPNTGGIGTAIFVAIGA AVMAFAVKGMKRR Majority									
510		520		530		540		550	
01	I	T	V	D	S	A	D	A	T
01	I	T	V	D	S	A	D	A	T
01	I	T	V	D	S	A	D	A	T
01	I	T	V	D	S	A	D	A	T
01	I	T	V	D	S	A	D	A	T
TKDN Majority									
51	T	K	D	N					
51	T	K	D	N					
51	T	K	D	N					
51	T	K	D	N					
51	T	K	D	N					

scoration 'Decoration #1': Shade (with solid black) residues that differ from the Consensus.

FIGURE 22A

134/487

M K K R Q K I W R G L S V T L L I L S Q I P F G I L V Q G E T Q D T N Q A L G K V I V K K T G D N A Majority																	
10		20		30		40		50									
1	M K K R Q K I W R G L S V T L L I L S Q I P F G I L V Q G E T Q D T N Q A L G K V I V K K T G D N A	sag649_2603.pep	1	M K K R Q K I W R G L S V T L L I L S Q I P F G I L V Q G E T Q D T N Q A L G K V I V K K T G D N A	sag649_coh1.pep	1	M K K R Q K I W R G L S V T L L I L S Q I P F G I L V Q G E T Q D T N Q A L G K V I V K K T G D N A	sag649_cjb111.pep	1	M K K R Q K I W R G L S V T L L I L S Q I P F G I L V Q G E T Q D T N Q A L G K V I V K K T G D N A	sag649_18rs21.pep	1	M K K R Q K I W R G L S V T L L I L S Q I P F G I L V Q G E T Q D T N Q A L G K V I V K K T G D N A	sag649_nem316.pep	1	M K K R Q K I W R G L S V T L L I L S Q I P F G I L V Q G E T Q D T N Q A L G K V I V K K T G D N A	sag649_cjb111.pep
T P L G K A T F V L K N D N D K S E T S H E T V E G S G E A T F E N I K P G D Y T L R E E T A P I G Majority																	
60		70		80		90		100									
51	T P L G K A T F V L K N D N D K S E T S H E T V E G S G E A T F E N I K P G D Y T L R E E T A P I G	sag649_2603.pep	51	T P L G K A T F V L K N D N D K S E T S H E T V E G S G E A T F E N I K P G D Y T L R E E T A P I G	sag649_coh1.pep	51	T P L G K A T F V L K N D N D K S E T S H E T V E G S G E A T F E N I K P G D Y T L R E E T A P I G	sag649_cjb111.pep	51	T P L G K A T F V L K N D N D K S E T S H E T V E G S G E A T F E N I K P G D Y T L R E E T A P I G	sag649_18rs21.pep	51	T P L G K A T F V L K N D N D K S E T S H E T V E G S G E A T F E N I K P G D Y T L R E E T A P I G	sag649_nem316.pep	51	T P L G K A T F V L K N D N D K S E T S H E T V E G S G E A T F E N I K P G D Y T L R E E T A P I G	sag649_cjb111.pep
Y K K T D K T W K V K V A D N G A T I I E G M D A D K A E K R K E V L N A Q Y P K S A I Y E D T K E Majority																	
110		120		130		140		150									
101	Y K K T D K T W K V K V A D N G A T I I E G M D A D K A E K R K E V L N A Q Y P K S A I Y E D T K E	sag649_2603.pep	101	Y K K T D K T W K V K V A D N G A T I I E G M D A D K A E K R K E V L N A Q Y P K S A I Y E D T K E	sag649_coh1.pep	101	Y K K T D K T W K V K V A D N G A T I I E G M D A D K A E K R K E V L N A Q Y P K S A I Y E D T K E	sag649_cjb111.pep	101	Y K K T D K T W K V K V A D N G A T I I E G M D A D K A E K R K E V L N A Q Y P K S A I Y E D T K E	sag649_18rs21.pep	101	Y K K T D K T W K V K V A D N G A T I I E G M D A D K A E K R K E V L N A Q Y P K S A I Y E D T K E	sag649_nem316.pep	101	Y K K T D K T W K V K V A D N G A T I I E G M D A D K A E K R K E V L N A Q Y P K S A I Y E D T K E	sag649_cjb111.pep
N Y P L V N V E G S K V G E Q Y K A L N P I N G K D G R R E I A E G W L S K K I T G V N D L D K N K Majority																	
160		170		180		190		200									
151	N Y P L V N V E G S K V G E Q Y K A L N P I N G K D G R R E I A E G W L S K K I T G V N D L D K N K	sag649_2603.pep	151	N Y P L V N V E G S K V G E Q Y K A L N P I N G K D G R R E I A E G W L S K K I T G V N D L D K N K	sag649_coh1.pep	151	N Y P L V N V E G S K V G E Q Y K A L N P I N G K D G R R E I A E G W L S K K I T G V N D L D K N K	sag649_cjb111.pep	151	N Y P L V N V E G S K V G E Q Y K A L N P I N G K D G R R E I A E G W L S K K I T G V N D L D K N K	sag649_18rs21.pep	151	N Y P L V N V E G S K V G E Q Y K A L N P I N G K D G R R E I A E G W L S K K I T G V N D L D K N K	sag649_nem316.pep	151	N Y P L V N V E G S K V G E Q Y K A L N P I N G K D G R R E I A E G W L S K K I T G V N D L D K N K	sag649_cjb111.pep
Y K I E L T V E G K T T V E T K E L N Q P L D V V V L L D N S N S M N N E R A N N S Q R A L K A G E Majority																	
210		220		230		240		250									
201	Y K I E L T V E G K T T V E T K E L N Q P L D V V V L L D N S N S M N N E R A N N S Q R A L K A G E	sag649_2603.pep	201	Y K I E L T V E G K T T V E T K E L N Q P L D V V V L L D N S N S M N N E R A N N S Q R A L K A G E	sag649_coh1.pep	201	Y K I E L T V E G K T T V E T K E L N Q P L D V V V L L D N S N S M N N E R A N N S Q R A L K A G E	sag649_cjb111.pep	201	Y K I E L T V E G K T T V E T K E L N Q P L D V V V L L D N S N S M N N E R A N N S Q R A L K A G E	sag649_18rs21.pep	201	Y K I E L T V E G K T T V E T K E L N Q P L D V V V L L D N S N S M N N E R A N N S Q R A L K A G E	sag649_nem316.pep	201	Y K I E L T V E G K T T V E T K E L N Q P L D V V V L L D N S N S M N N E R A N N S Q R A L K A G E	sag649_cjb111.pep
A V E K L I D K I T S N K D N R V A L V T Y A S T I F D G T E A T V S K G V A D Q N G K A L N D S V Majority																	
260		270		280		290		300									
251	A V E K L I D K I T S N K D N R V A L V T Y A S T I F D G T E A T V S K G V A D Q N G K A L N D S V	sag649_2603.pep	251	A V E K L I D K I T S N K D N R V A L V T Y A S T I F D G T E A T V S K G V A D Q N G K A L N D S V	sag649_coh1.pep	251	A V E K L I D K I T S N K D N R V A L V T Y A S T I F D G T E A T V S K G V A D Q N G K A L N D S V	sag649_cjb111.pep	251	A V E K L I D K I T S N K D N R V A L V T Y A S T I F D G T E A T V S K G V A D Q N G K A L N D S V	sag649_18rs21.pep	251	A V E K L I D K I T S N K D N R V A L V T Y A S T I F D G T E A T V S K G V A D Q N G K A L N D S V	sag649_nem316.pep	251	A V E K L I D K I T S N K D N R V A L V T Y A S T I F D G T E A T V S K G V A D Q N G K A L N D S V	sag649_cjb111.pep
S W D Y H K T T F T A T T H N Y S Y L N L T N D A N E V N I L K S R I P K E A E H I N G D R T L Y Q Majority																	
310		320		330		340		350									
301	S W D Y H K T T F T A T T H N Y S Y L N L T N D A N E V N I L K S R I P K E A E H I N G D R T L Y Q	sag649_2603.pep	301	S W D Y H K T T F T A T T H N Y S Y L N L T N D A N E V N I L K S R I P K E A E H I N G D R T L Y Q	sag649_coh1.pep	301	S W D Y H K T T F T A T T H N Y S Y L N L T N D A N E V N I L K S R I P K E A E H I N G D R T L Y Q	sag649_cjb111.pep	301	S W D Y H K T T F T A T T H N Y S Y L N L T N D A N E V N I L K S R I P K E A E H I N G D R T L Y Q	sag649_18rs21.pep	301	S W D Y H K T T F T A T T H N Y S Y L N L T N D A N E V N I L K S R I P K E A E H I N G D R T L Y Q	sag649_nem316.pep	301	S W D Y H K T T F T A T T H N Y S Y L N L T N D A N E V N I L K S R I P K E A E H I N G D R T L Y Q	sag649_cjb111.pep
F G A T F T Q K A L M K A N E I L E T Q S S N A R K K L I F H V T D G V P T M S Y A I N F N P Y I S Majority																	
360		370		380		390		400									
351	F G A T F T Q K A L M K A N E I L E T Q S S N A R K K L I F H V T D G V P T M S Y A I N F N P Y I S	sag649_2603.pep	351	F G A T F T Q K A L M K A N E I L E T Q S S N A R K K L I F H V T D G V P T M S Y A I N F N P Y I S	sag649_coh1.pep	351	F G A T F T Q K A L M K A N E I L E T Q S S N A R K K L I F H V T D G V P T M S Y A I N F N P Y I S	sag649_cjb111.pep	351	F G A T F T Q K A L M K A N E I L E T Q S S N A R K K L I F H V T D G V P T M S Y A I N F N P Y I S	sag649_18rs21.pep	351	F G A T F T Q K A L M K A N E I L E T Q S S N A R K K L I F H V T D G V P T M S Y A I N F N P Y I S	sag649_nem316.pep	351	F G A T F T Q K A L M K A N E I L E T Q S S N A R K K L I F H V T D G V P T M S Y A I N F N P Y I S	sag649_cjb111.pep

Figure 23

135/487

TSYQNQFNSFLNKKIPDRSGILQEDFIINGDDYQIVKGDGESFKLFSDRKV Majority									
410		420		430		440		450	
401	TSYQNQFNSFLNKKIPDRSGILQEDFIINGDDYQIVKGDGESFKLFSDRKV	sag649_2603.pep							
401	TSYQNQFNSFLNKKIPDRSGILQEDFIINGDDYQIVKGDGESFKLFSDRKV	sag649_cohl.pep							
401	TSYQNQFNSFLNKKIPDRSGILQEDFIINGDDYQIVKGDGESFKLFSDRKV	sag649_cjb111.pep							
401	TSYQNQFNSFLNKKIPDRSGILQEDFIINGDDYQIVKGDGESFKLFSDRKV	sag649_18rs21.pep							
401	TSYQNQFNSFLNKKIPDRSGILQEDFIINGDDYQIVKGDGESFKLFSDRKV	sag649_nem316.pep							
401	TSYQNQFNSFLNKKIPDRSGILQEDFIINGDDYQIVKGDGESFKLFSDRKV	sag649_cjb111.pep							
PVTGCTTQAAAYRVPQNQLSVMSNEGYAINSGYIYLYWRDYNWVYFPDPKT Majority									
460		470		480		490		500	
451	PVTGCTTQAAAYRVPQNQLSVMSNEGYAINSGYIYLYWRDYNWVYFPDPKT	sag649_2603.pep							
451	PVTGCTTQAAAYRVPQNQLSVMSNEGYAINSGYIYLYWRDYNWVYFPDPKT	sag649_cohl.pep							
451	PVTGCTTQAAAYRVPQNQLSVMSNEGYAINSGYIYLYWRDYNWVYFPDPKT	sag649_cjb111.pep							
451	PVTGCTTQAAAYRVPQNQLSVMSNEGYAINSGYIYLYWRDYNWVYFPDPKT	sag649_18rs21.pep							
451	PVTGCTTQAAAYRVPQNQLSVMSNEGYAINSGYIYLYWRDYNWVYFPDPKT	sag649_nem316.pep							
451	PVTGCTTQAAAYRVPQNQLSVMSNEGYAINSGYIYLYWRDYNWVYFPDPKT	sag649_cjb111.pep							
KKVSATKQIKITHGEPTTLYFNGNIRPKGYDIFTVIGVNGDPCATPLEAE Majority									
510		520		530		540		550	
501	KKVSATKQIKITHGEPTTLYFNGNIRPKGYDIFTVIGVNGDPCATPLEAE	sag649_2603.pep							
501	KKVSATKQIKITHGEPTTLYFNGNIRPKGYDIFTVIGVNGDPCATPLEAE	sag649_cohl.pep							
501	KKVSATKQIKITHGEPTTLYFNGNIRPKGYDIFTVIGVNGDPCATPLEAE	sag649_cjb111.pep							
501	KKVSATKQIKITHGEPTTLYFNGNIRPKGYDIFTVIGVNGDPCATPLEAE	sag649_18rs21.pep							
501	KKVSATKQIKITHGEPTTLYFNGNIRPKGYDIFTVIGVNGDPCATPLEAE	sag649_nem316.pep							
501	KKVSATKQIKITHGEPTTLYFNGNIRPKGYDIFTVIGVNGDPCATPLEAE	sag649_cjb111.pep							
KFMQSISSTKENYTNVDDTNKIYDELNKFYKTIIVEEKHSIVDGNVTDPMG Majority									
560		570		580		590		600	
551	KFMQSISSTKENYTNVDDTNKIYDELNKFYKTIIVEEKHSIVDGNVTDPMG	sag649_2603.pep							
551	KFMQSISSTKENYTNVDDTNKIYDELNKFYKTIIVEEKHSIVDGNVTDPMG	sag649_cohl.pep							
551	KFMQSISSTKENYTNVDDTNKIYDELNKFYKTIIVEEKHSIVDGNVTDPMG	sag649_cjb111.pep							
551	KFMQSISSTKENYTNVDDTNKIYDELNKFYKTIIVEEKHSIVDGNVTDPMG	sag649_18rs21.pep							
551	KFMQSISSTKENYTNVDDTNKIYDELNKFYKTIIVEEKHSIVDGNVTDPMG	sag649_nem316.pep							
551	KFMQSISSTKENYTNVDDTNKIYDELNKFYKTIIVEEKHSIVDGNVTDPMG	sag649_cjb111.pep							
EMIEFQLKNGQSFTHDDYVLVGNDCGSQLKNGVALGCPNSDGGILKDVTVT Majority									
610		620		630		640		650	
601	EMIEFQLKNGQSFTHDDYVLVGNDCGSQLKNGVALGCPNSDGGILKDVTVT	sag649_2603.pep							
601	EMIEFQLKNGQSFTHDDYVLVGNDCGSQLKNGVALGCPNSDGGILKDVTVT	sag649_cohl.pep							
601	EMIEFQLKNGQSFTHDDYVLVGNDCGSQLKNGVALGCPNSDGGILKDVTVT	sag649_cjb111.pep							
601	EMIEFQLKNGQSFTHDDYVLVGNDCGSQLKNGVALGCPNSDGGILKDVTVT	sag649_18rs21.pep							
601	EMIEFQLKNGQSFTHDDYVLVGNDCGSQLKNGVALGCPNSDGGILKDVTVT	sag649_nem316.pep							
601	EMIEFQLKNGQSFTHDDYVLVGNDCGSQLKNGVALGCPNSDGGILKDVTVT	sag649_cjb111.pep							
YDKTSQTIKINHLNLGSGQKVVLTYDVRCLKDNYISNKFYNTNNRRTTSLSPK Majority									
660		670		680		690		700	
651	YDKTSQTIKINHLNLGSGQKVVLTYDVRCLKDNYISNKFYNTNNRRTTSLSPK	sag649_2603.pep							
651	YDKTSQTIKINHLNLGSGQKVVLTYDVRCLKDNYISNKFYNTNNRRTTSLSPK	sag649_cohl.pep							
651	YDKTSQTIKINHLNLGSGQKVVLTYDVRCLKDNYISNKFYNTNNRRTTSLSPK	sag649_cjb111.pep							
651	YDKTSQTIKINHLNLGSGQKVVLTYDVRCLKDNYISNKFYNTNNRRTTSLSPK	sag649_18rs21.pep							
651	YDKTSQTIKINHLNLGSGQKVVLTYDVRCLKDNYISNKFYNTNNRRTTSLSPK	sag649_nem316.pep							
651	YDKTSQTIKINHLNLGSGQKVVLTYDVRCLKDNYISNKFYNTNNRRTTSLSPK	sag649_cjb111.pep							
SEKEPNTIRDFPIPKIRDVREFPVLTIISNQKKMGVEFEIKVKNKDKHSES L Majority									
710		720		730		740		750	
701	SEKEPNTIRDFPIPKIRDVREFPVLTIISNQKKMGVEFEIKVKNKDKHSES L	sag649_2603.pep							
701	SEKEPNTIRDFPIPKIRDVREFPVLTIISNQKKMGVEFEIKVKNKDKHSES L	sag649_cohl.pep							
701	SEKEPNTIRDFPIPKIRDVREFPVLTIISNQKKMGVEFEIKVKNKDKHSES L	sag649_cjb111.pep							
701	SEKEPNTIRDFPIPKIRDVREFPVLTIISNQKKMGVEFEIKVKNKDKHSES L	sag649_18rs21.pep							
701	SEKEPNTIRDFPIPKIRDVREFPVLTIISNQKKMGVEFEIKVKNKDKHSES L	sag649_nem316.pep							
701	SEKEPNTIRDFPIPKIRDVREFPVLTIISNQKKMGVEFEIKVKNKDKHSES L	sag649_cjb111.pep							
LGAKFQLQIEKDFSGYKQFVPEGSDVTTKNDGKIYFKALQDGNKYKLYEIS Majority									
760		770		780		790		800	
751	LGAKFQLQIEKDFSGYKQFVPEGSDVTTKNDGKIYFKALQDGNKYKLYEIS	sag649_2603.pep							
751	LGAKFQLQIEKDFSGYKQFVPEGSDVTTKNDGKIYFKALQDGNKYKLYEIS	sag649_cohl.pep							
751	LGAKFQLQIEKDFSGYKQFVPEGSDVTTKNDGKIYFKALQDGNKYKLYEIS	sag649_cjb111.pep							
751	LGAKFQLQIEKDFSGYKQFVPEGSDVTTKNDGKIYFKALQDGNKYKLYEIS	sag649_18rs21.pep							
751	LGAKFQLQIEKDFSGYKQFVPEGSDVTTKNDGKIYFKALQDGNKYKLYEIS	sag649_nem316.pep							
751	LGAKFQLQIEKDFSGYKQFVPEGSDVTTKNDGKIYFKALQDGNKYKLYEIS	sag649_cjb111.pep							

FIGURE 23A

136/487

SPDGYIEVKTKPVTFTIQNGEVTNLKADPNANKNQIGYLEGNGKHLITN																																								Majority																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																															
810										820										830										840										850																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																															
801	S	P	D	G	Y	I	E	V	K	T	K	P	V	T	F	T	I	Q	N	G	E	V	T	N	L	K	A	D	P	N	A	N	K	N	Q	I	G	Y	L	E	G	N	G	K	H	L	I	T	N		sag649_2603.pep																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
801	S	P	D	G	Y	I	E	V	K	T	K	P	V	T	F	T	I	Q	N	G	E	V	T	N	L	K	A	D	P	N	A	N	K	N	Q	I	G	Y	L	E	G	N	G	K	H	L	I	T	N		sag649_coh1.pep																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
801	S	P	D	G	Y	I	E	V	K	T	K	P	V	T	F	T	I	Q	N	G	E	V	T	N	L	K	A	D	P	N	A	N	K	N	Q	I	G	Y	L	E	G	N	G	K	H	L	I	T	N		sag649_cjb111.pep																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
801	S	P	D	G	Y	I	E	V	K	T	K	P	V	T	F	T	I	Q	N	G	E	V	T	N	L	K	A	D	P	N	A	N	K	N	Q	I	G	Y	L	E	G	N	G	K	H	L	I	T	N		sag649_18rs21.pep																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
801	S	P	D	G	Y	I	E	V	K	T	K	P	V	T	F	T	I	Q	N	G	E	V	T	N	L	K	A	D	P	N	A	N	K	N	Q	I	G	Y	L	E	G	N	G	K	H	L	I	T	N		sag649_new316.pep																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
801	S	P	D	G	Y	I	E	V	K	T	K	P	V	T	F	T	I	Q	N	G	E	V	T	N	L	K	A	D	P	N	A	N	K	N	Q	I	G	Y	L	E	G	N	G	K	H	L	I	T	N		sag649_cjb111.pep																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
TPKRPPGCVFPKTCGGIGTIVYILVGSTFMILTICSFRRKQL																																								Majority																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																															
860										870										880										890																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
851	T	P	K	R	P	P	G	V	F	P	K	T	C	G	G	I	G	T	I	V	Y	I	L	V	G	S	T	F	M	I	L	T	I	C	S	F	R	R	K	Q	L																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														</



Figure 24

138/487

SYAINNFKLGCASYESQFEQMKKNGYLNKSNFLLTDKPEDIKNGGESYFLF Majority									
410 420 430 440 450									
401	SYAINNFKLGCASYESQFEQMKKNGYLNKSNFLLTDKPEDIKNGGESYFLF								
401	SYAINNFKLGCASYESQFEQMKKNGYLNKSNFLLTDKPEDIKNGGESYFLF								sag1408_2603.pep
401	SYAINNFKLGCASYESQFEQMKKNGYLNKSNFLLTDKPEDIKNGGESYFLF								sag1408_515.pep
401	SYAINNFKLGCASYESQFEQMKKNGYLNKSNFLLTDKPEDIKNGGESYFLF								sag1408_18rs21.seq
396	SYAINNFKLGCASYESQFEQMKKNGYLNKSNFLLTDKPEDIKNGGESYFLF								sag1408_cjb111.pep
401	SYAINNFKLGCASYESQFEQMKKNGYLNKSNFLLTDKPEDIKNGGESYFLF								sag1408_h36b.pep
401	SYAINNFKLGCASYESQFEQMKKNGYLNKSNFLLTDKPEDIKNGGESYFLF								sag1408_nem316.pep
PLDSYQTQIIISGNLQKLHYLDLNLNYPKGTIYRNGPVEHGTPTKLYINS Majority									
460 470 480 490 500									
451	PLDSYQTQIIISGNLQKLHYLDLNLNYPKGTIYRNGPVEHGTPTKLYINS								
451	PLDSYQTQIIISGNLQKLHYLDLNLNYPKGTIYRNGPVEHGTPTKLYINS								sag1408_2603.pep
451	PLDSYQTQIIISGNLQKLHYLDLNLNYPKGTIYRNGPVEHGTPTKLYINS								sag1408_515.pep
451	PLDSYQTQIIISGNLQKLHYLDLNLNYPKGTIYRNGPVEHGTPTKLYINS								sag1408_18rs21.seq
451	PLDSYQTQIIISGNLQKLHYLDLNLNYPKGTIYRNGPVEHGTPTKLYINS								sag1408_cjb111.pep
446	PLDSYQTQIIISGNLQKLHYLDLNLNYPKGTIYRNGPVEHGTPTKLYINS								sag1408_h36b.pep
451	PLDSYQTQIIISGNLQKLHYLDLNLNYPKGTIYRNGPVEHGTPTKLYINS								sag1408_nem316.pep
LKQKNYDIFNFGIDISGFRQVYNEFYKKNQDGTFFQKLKEEAFELSDGEIT Majority									
510 520 530 540 550									
501	LKQKNYDIFNFGIDISGFRQVYNEFYKKNQDGTFFQKLKEEAFELSDGEIT								
501	LKQKNYDIFNFGIDISGFRQVYNEFYKKNQDGTFFQKLKEEAFELSDGEIT								sag1408_2603.pep
501	LKQKNYDIFNFGIDISGFRQVYNEFYKKNQDGTFFQKLKEEAFELSDGEIT								sag1408_515.pep
501	LKQKNYDIFNFGIDISGFRQVYNEFYKKNQDGTFFQKLKEEAFELSDGEIT								sag1408_18rs21.seq
496	LKQKNYDIFNFGIDISGFRQVYNEFYKKNQDGTFFQKLKEEAFELSDGEIT								sag1408_cjb111.pep
501	LKQKNYDIFNFGIDISGFRQVYNEFYKKNQDGTFFQKLKEEAFELSDGEIT								sag1408_h36b.pep
501	LKQKNYDIFNFGIDISGFRQVYNEFYKKNQDGTFFQKLKEEAFELSDGEIT								sag1408_nem316.pep
ELMRSFSSSKPEYYTPIVTSADTSNNEILSKIQQQFETILTKEINSIVNGTI Majority									
560 570 580 590 600									
551	ELMRSFSSSKPEYYTPIVTSADTSNNEILSKIQQQFETILTKEINSIVNGTI								
551	ELMRSFSSSKPEYYTPIVTSADTSNNEILSKIQQQFETILTKEINSIVNGTI								sag1408_2603.pep
551	ELMRSFSSSKPEYYTPIVTSADTSNNEILSKIQQQFETILTKEINSIVNGTI								sag1408_515.pep
551	ELMRSFSSSKPEYYTPIVTSADTSNNEILSKIQQQFETILTKEINSIVNGTI								sag1408_18rs21.seq
546	ELMRSFSSSKPEYYTPIVTSADTSNNEILSKIQQQFETILTKEINSIVNGTI								sag1408_cjb111.pep
551	ELMRSFSSSKPEYYTPIVTSADTSNNEILSKIQQQFETILTKEINSIVNGTI								sag1408_h36b.pep
551	ELMRSFSSSKPEYYTPIVTSADTSNNEILSKIQQQFETILTKEINSIVNGTI								sag1408_nem316.pep
EDPMGDKINLQLGNGQTLQPSDYTLQGNDSVMKDGIAATGCPNNDGGILK Majority									
610 620 630 640 650									
601	EDPMGDKINLQLGNGQTLQPSDYTLQGNDSVMKDGIAATGCPNNDGGILK								
601	EDPMGDKINLQLGNGQTLQPSDYTLQGNDSVMKDGIAATGCPNNDGGILK								sag1408_2603.pep
601	EDPMGDKINLQLGNGQTLQPSDYTLQGNDSVMKDGIAATGCPNNDGGILK								sag1408_515.pep
601	EDPMGDKINLQLGNGQTLQPSDYTLQGNDSVMKDGIAATGCPNNDGGILK								sag1408_18rs21.seq
596	EDPMGDKINLQLGNGQTLQPSDYTLQGNDSVMKDGIAATGCPNNDGGILK								sag1408_cjb111.pep
601	EDPMGDKINLQLGNGQTLQPSDYTLQGNDSVMKDGIAATGCPNNDGGILK								sag1408_h36b.pep
601	EDPMGDKINLQLGNGQTLQPSDYTLQGNDSVMKDGIAATGCPNNDGGILK								sag1408_nem316.pep
GVKLEYIGNKLYVRGLNLGEGQKVTLTYDVKLDDSFISNKFYDTNGRTTL Majority									
660 670 680 690 700									
651	GVKLEYIGNKLYVRGLNLGEGQKVTLTYDVKLDDSFISNKFYDTNGRTTL								
651	GVKLEYIGNKLYVRGLNLGEGQKVTLTYDVKLDDSFISNKFYDTNGRTTL								sag1408_2603.pep
651	GVKLEYIGNKLYVRGLNLGEGQKVTLTYDVKLDDSFISNKFYDTNGRTTL								sag1408_515.pep
651	GVKLEYIGNKLYVRGLNLGEGQKVTLTYDVKLDDSFISNKFYDTNGRTTL								sag1408_18rs21.seq
646	GVKLEYIGNKLYVRGLNLGEGQKVTLTYDVKLDDSFISNKFYDTNGRTTL								sag1408_cjb111.pep
651	GVKLEYIGNKLYVRGLNLGEGQKVTLTYDVKLDDSFISNKFYDTNGRTTL								sag1408_h36b.pep
651	GVKLEYIGNKLYVRGLNLGEGQKVTLTYDVKLDDSFISNKFYDTNGRTTL								sag1408_nem316.pep
NPKSEDPNTLRDFFIPKIRDVREYPTITIKNEKKLGEIEFIKVDKDNKKL Majority									
710 720 730 740 750									
701	NPKSEDPNTLRDFFIPKIRDVREYPTITIKNEKKLGEIEFIKVDKDNKKL								
701	NPKSEDPNTLRDFFIPKIRDVREYPTITIKNEKKLGEIEFIKVDKDNKKL								sag1408_2603.pep
701	NPKSEDPNTLRDFFIPKIRDVREYPTITIKNEKKLGEIEFIKVDKDNKKL								sag1408_515.pep
701	NPKSEDPNTLRDFFIPKIRDVREYPTITIKNEKKLGEIEFIKVDKDNKKL								sag1408_18rs21.seq
696	NPKSEDPNTLRDFFIPKIRDVREYPTITIKNEKKLGEIEFIKVDKDNKKL								sag1408_cjb111.pep
701	NPKSEDPNTLRDFFIPKIRDVREYPTITIKNEKKLGEIEFIKVDKDNKKL								sag1408_h36b.pep
701	NPKSEDPNTLRDFFIPKIRDVREYPTITIKNEKKLGEIEFIKVDKDNKKL								sag1408_nem316.pep
LLKGATFELQEFNEDYKLYLPIKNNNSKVVTGENGKISYKDLKDGKYQLI Majority									
760 770 780 790 800									
751	LLKGATFELQEFNEDYKLYLPIKNNNSKVVTGENGKISYKDLKDGKYQLI								
751	LLKGATFELQEFNEDYKLYLPIKNNNSKVVTGENGKISYKDLKDGKYQLI								sag1408_2603.pep
751	LLKGATFELQEFNEDYKLYLPIKNNNSKVVTGENGKISYKDLKDGKYQLI								sag1408_515.pep
751	LLKGATFELQEFNEDYKLYLPIKNNNSKVVTGENGKISYKDLKDGKYQLI								sag1408_18rs21.seq
746	LLKGATFELQEFNEDYKLYLPIKNNNSKVVTGENGKISYKDLKDGKYQLI								sag1408_cjb111.pep
751	LLKGATFELQEFNEDYKLYLPIKNNNSKVVTGENGKISYKDLKDGKYQLI								sag1408_h36b.pep
751	LLKGATFELQEFNEDYKLYLPIKNNNSKVVTGENGKISYKDLKDGKYQLI								sag1408_nem316.pep

FIGURE 24A

139/487

E A V S P E D Y Q K I T N K P I L T F E V V K G S I Q N I I A V N K Q I S E Y H E E G D K H L I T N Majority										
	810		820		830		840		850	
801	E A V S P E D Y Q K I T N K P I L T F E V V K G S I Q N I I A V N K Q I S E Y H E E G D K H L I T N	sag1408_2603.pep								
801	E A V S P E D Y Q K I T N K P I L T F E V V K G S I Q N I I A V N K Q I S E Y H E E G D K H L I T N	sag1408_515.pep								
801	E A V S P E D Y Q K I T N K P I L T F E V V K G S I Q N I I A V N K Q I S E Y H E E G D K H L I T N	sag1408_18rs21.seq								
796	E A V S P E D Y Q K I T N K P I L T F E V V K G S I Q N I I A V N K Q I S E Y H E E G D K H L I T N	sag1408_cjb111.pep								
801	E A V S P E D Y Q K I T N K P I L T F E V V K G S I Q N I I A V N K Q I S E Y H E E G D K H L I T N	sag1408_h36b.pep								
801	E A V S P E D Y Q K I T N K P I L T F E V V K G S I Q N I I A V N K Q I S E Y H E E G D K H L I T N	sag1408_nem316.pep								

T H I P P K G I I P M T G G K G I L S F I L I G G A M M S I A G G I Y I W K R Y K K S S D M S I E K Majority										
	860		870		880		890		900	
851	T H I P P K G I I P M T G G K G I L S F I L I G G A M M S I A G G I Y I W K R Y K K S S D M S I E K	sag1408_2603.pep								
851	T H I P P K G I I P M T G G K G I L S F I L I G G A M M S I A G G I Y I W K R Y K K S S D M S I E K	sag1408_515.pep								
851	T H I P P K G I I P M T G G K G I L S F I L I G G A M M S I A G G I Y I W K R Y K K S S D M S I E K	sag1408_18rs21.seq								
851	T H I P P K G I I P M T G G K G I L S F I L I G G A M M S I A G G I Y I W K R Y K K S S D M S I E K	sag1408_cjb111.pep								
846	T H I P P K G I I P M T G G K G I L S F I L I G G A M M S I A G G I Y I W K R Y K K S S D M S I E K	sag1408_h36b.pep								
851	T H I P P K G I I P M T G G K G I L S F I L I G G A M M S I A G G I Y I W K R Y K K S S D M S I E K	sag1408_nem316.pep								

D

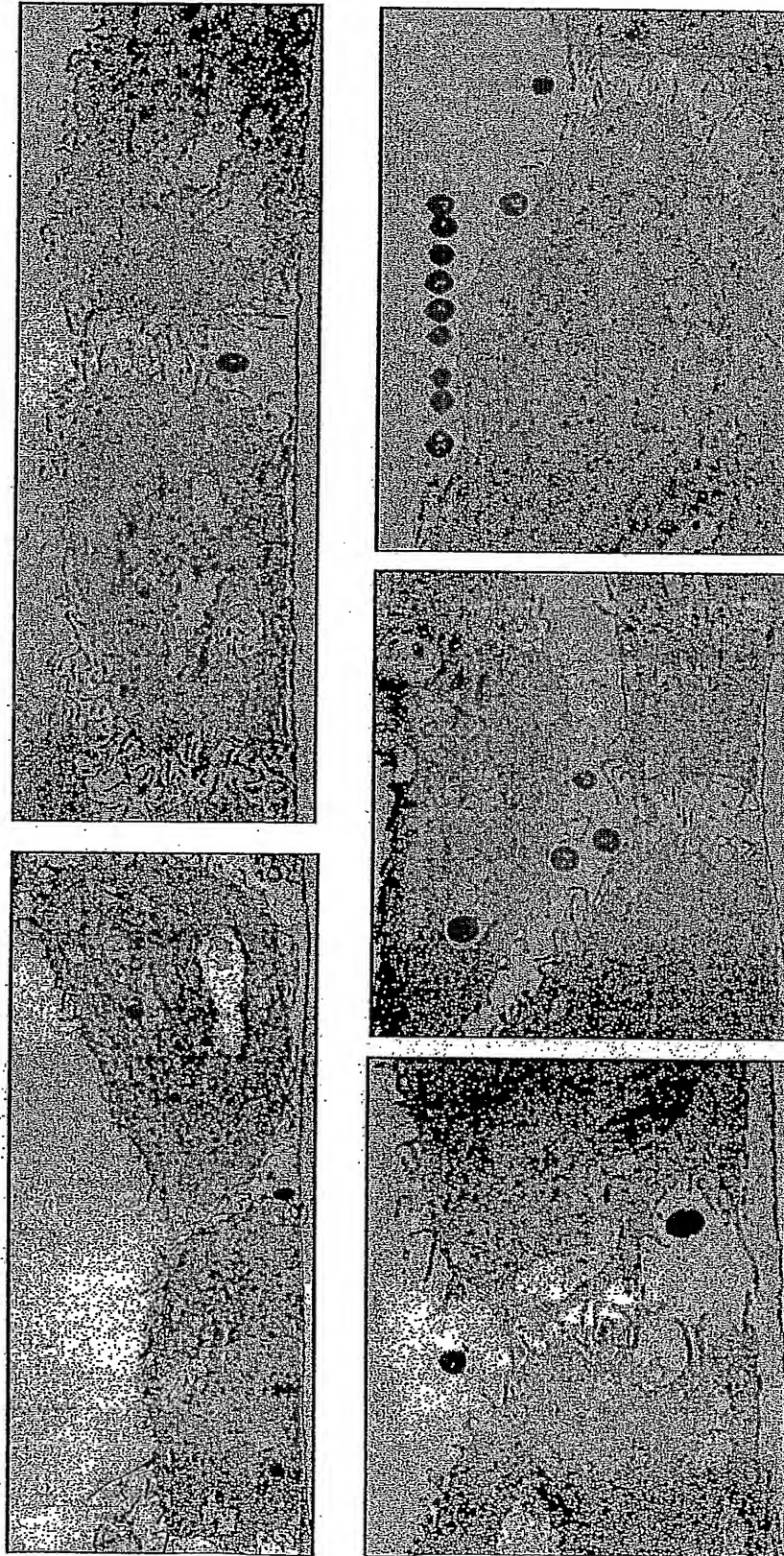
Majority

901 D	sag1408_2603.pep
901 D	sag1408_515.pep
901 D	sag1408_18rs21.seq
901 D	sag1408_cjb111.pep
896 D	sag1408_h36b.pep
901 D	sag1408_nem316.pep

Decoration 'Decoration #1': Shade (with solid black) residues that differ from the Consensus.

FIGURE 24B

Figure 25: GBS closely associate with tight junctions and cross the monolayer by a paracellular route



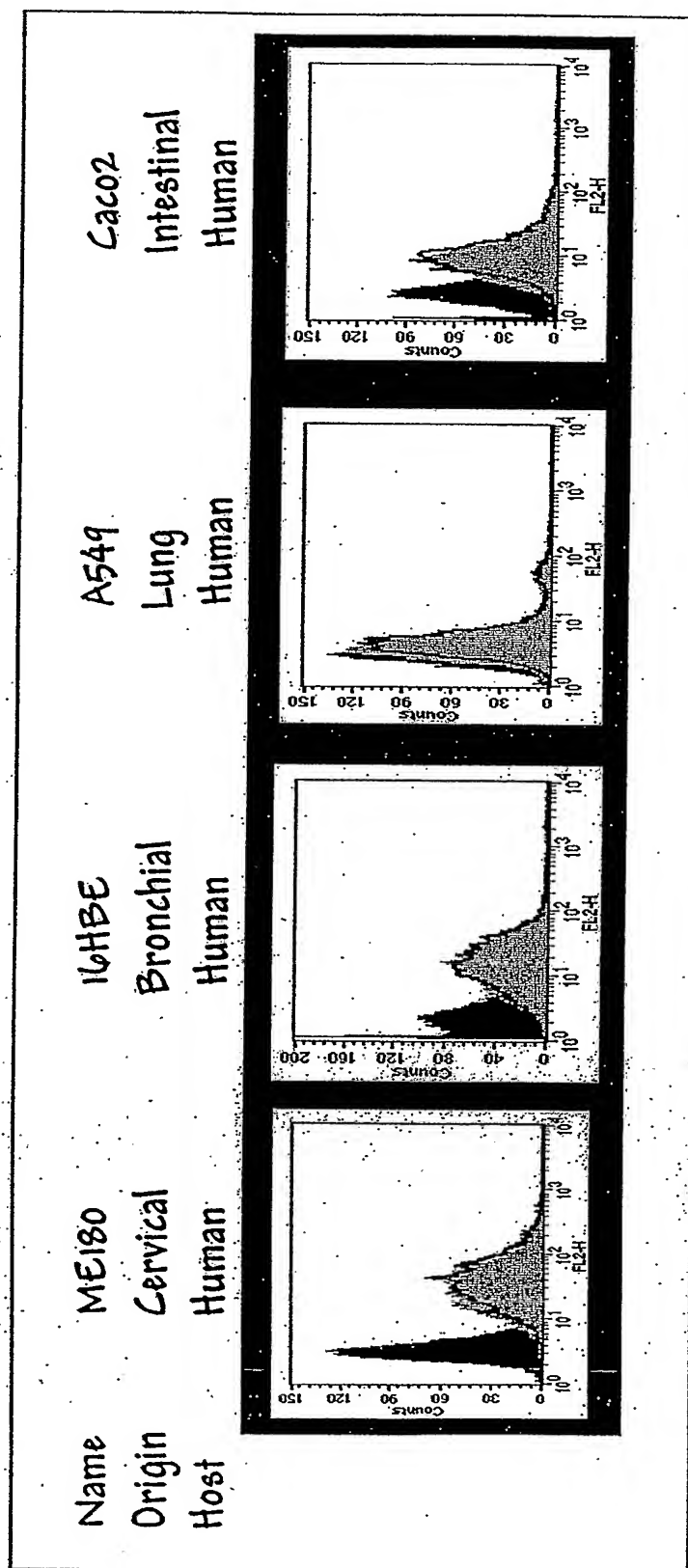
Transmission Electron Microscopy images of GBS infection  
of ME180 cervical epithelial cells.

Figure 26: GBS infection of ME180 cells  
GBS infection of ME180 cells



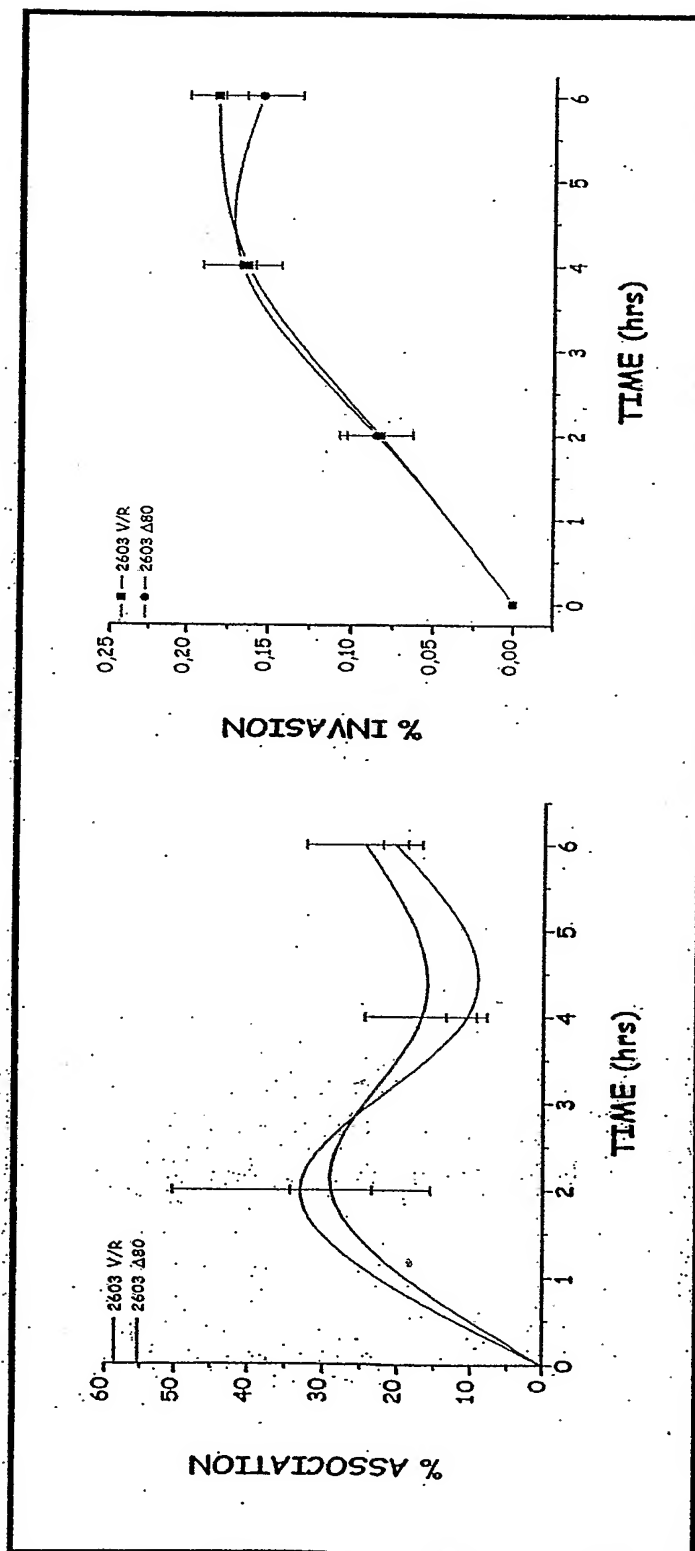
PCT/US05/27239/42/487

Figure 27



PCT/US05/27239 143/487

Figure 28





144/487

Figure 29

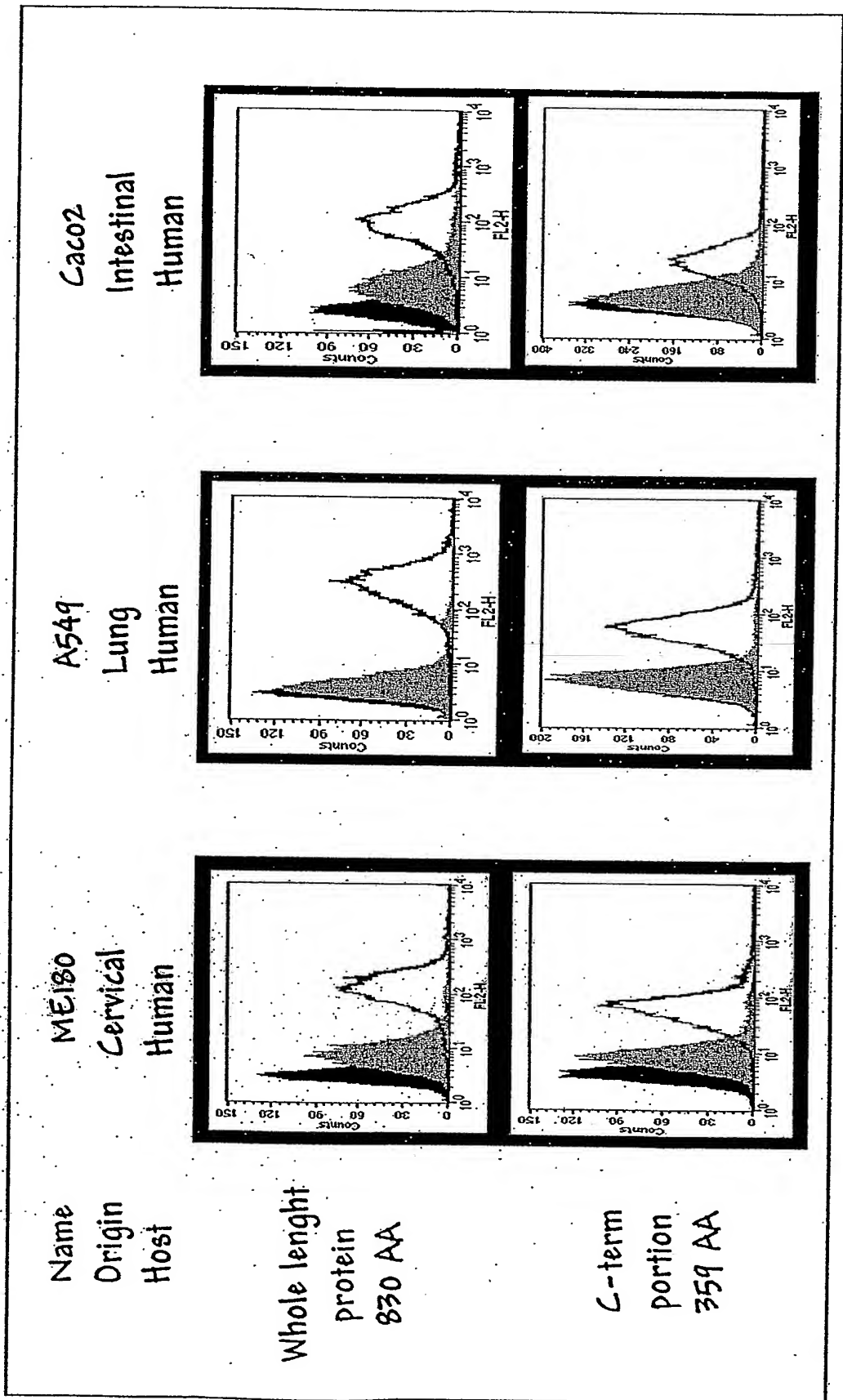
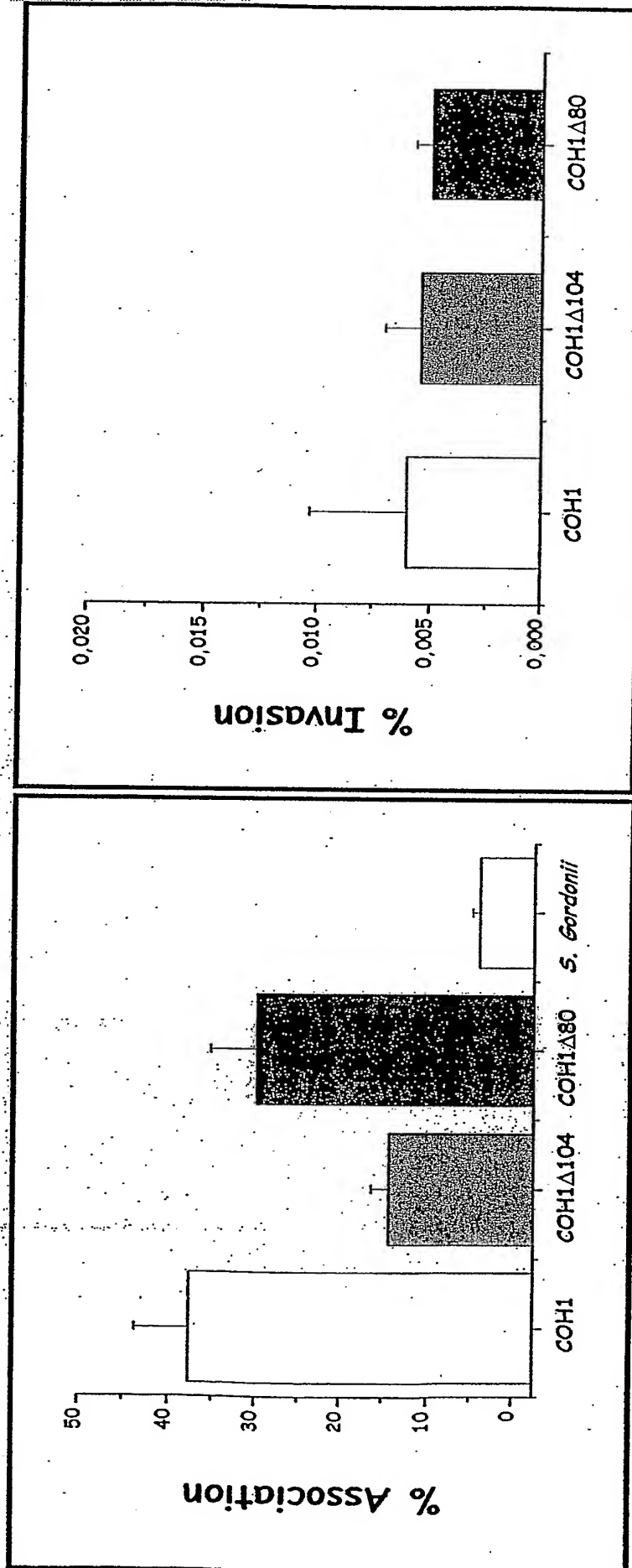
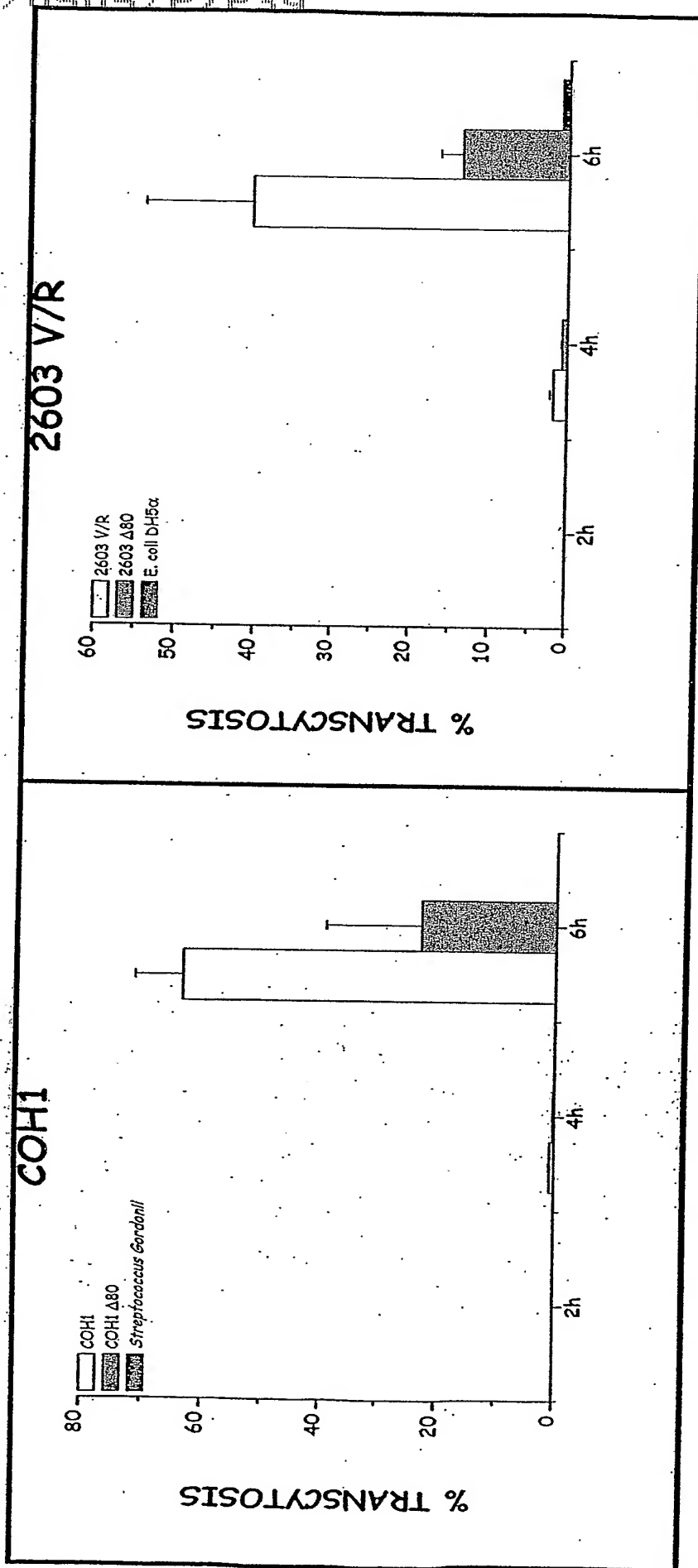


Figure 30



PCT/US2005/027239

Figure 31



PCT/US2005/027239/47/487

Figure 32

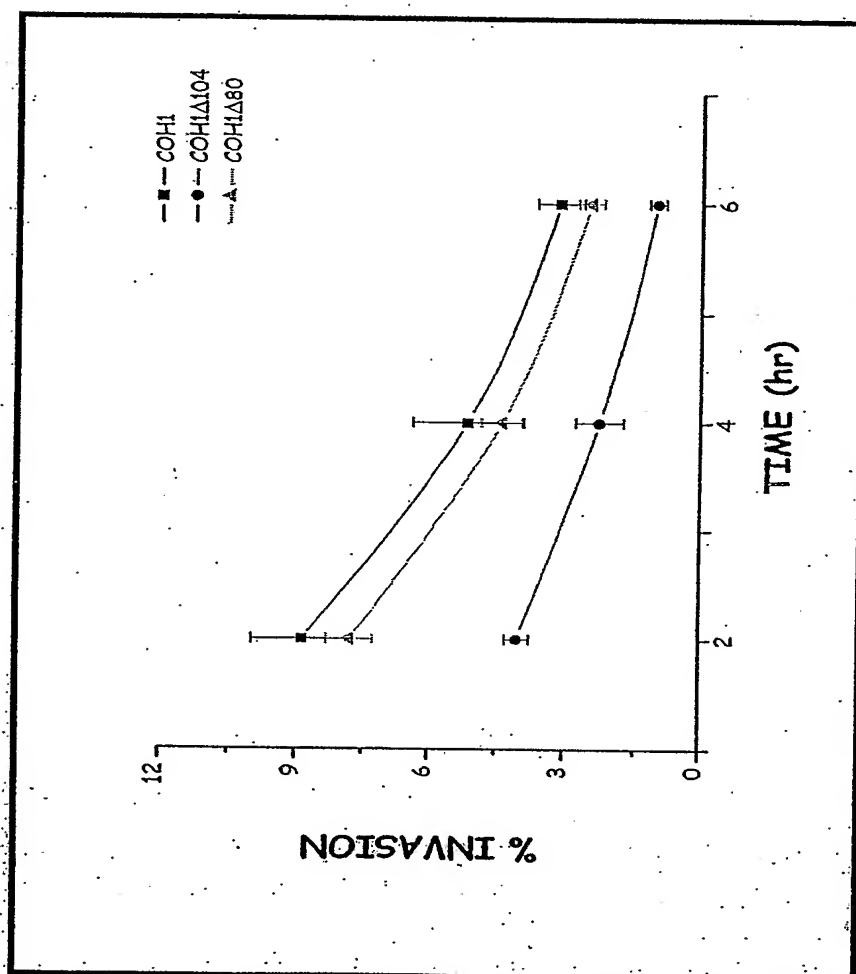
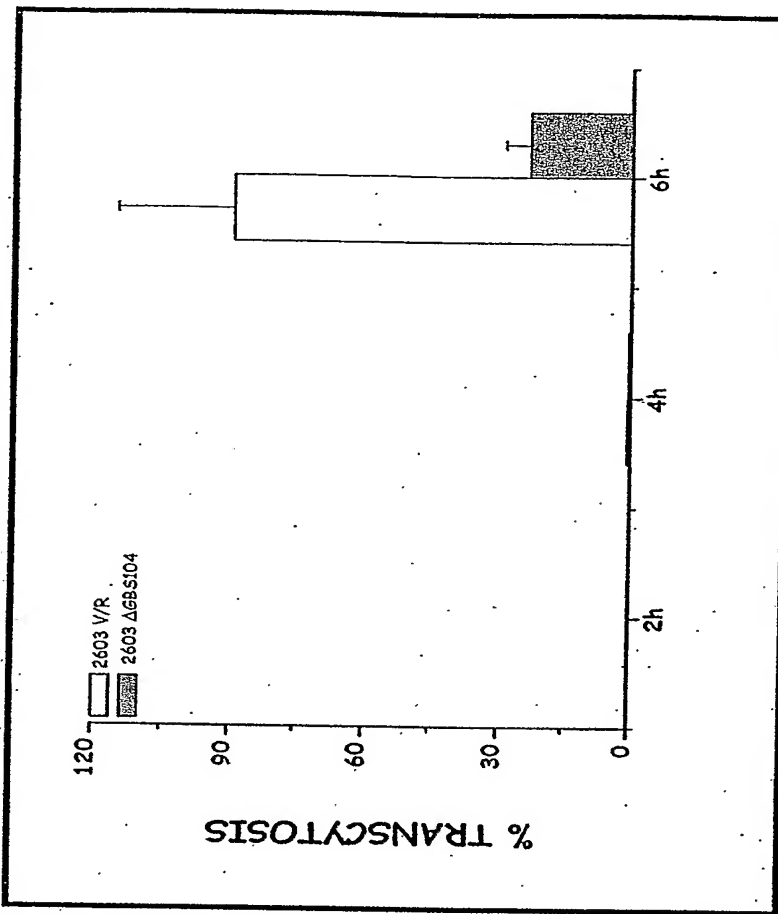
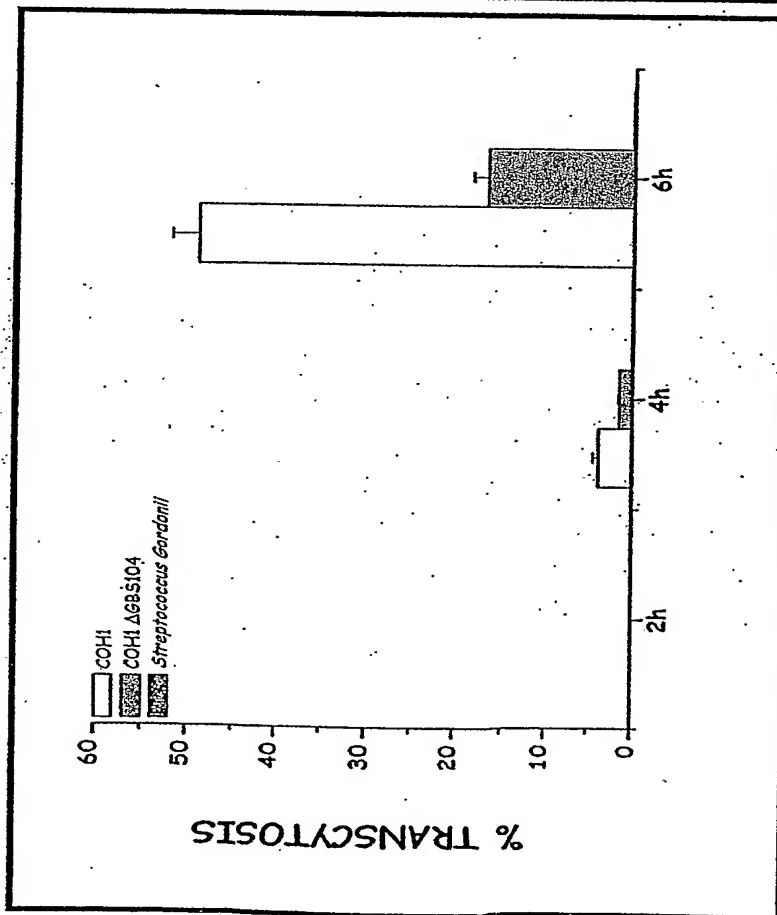


Figure 33

COH1

2603 V/R

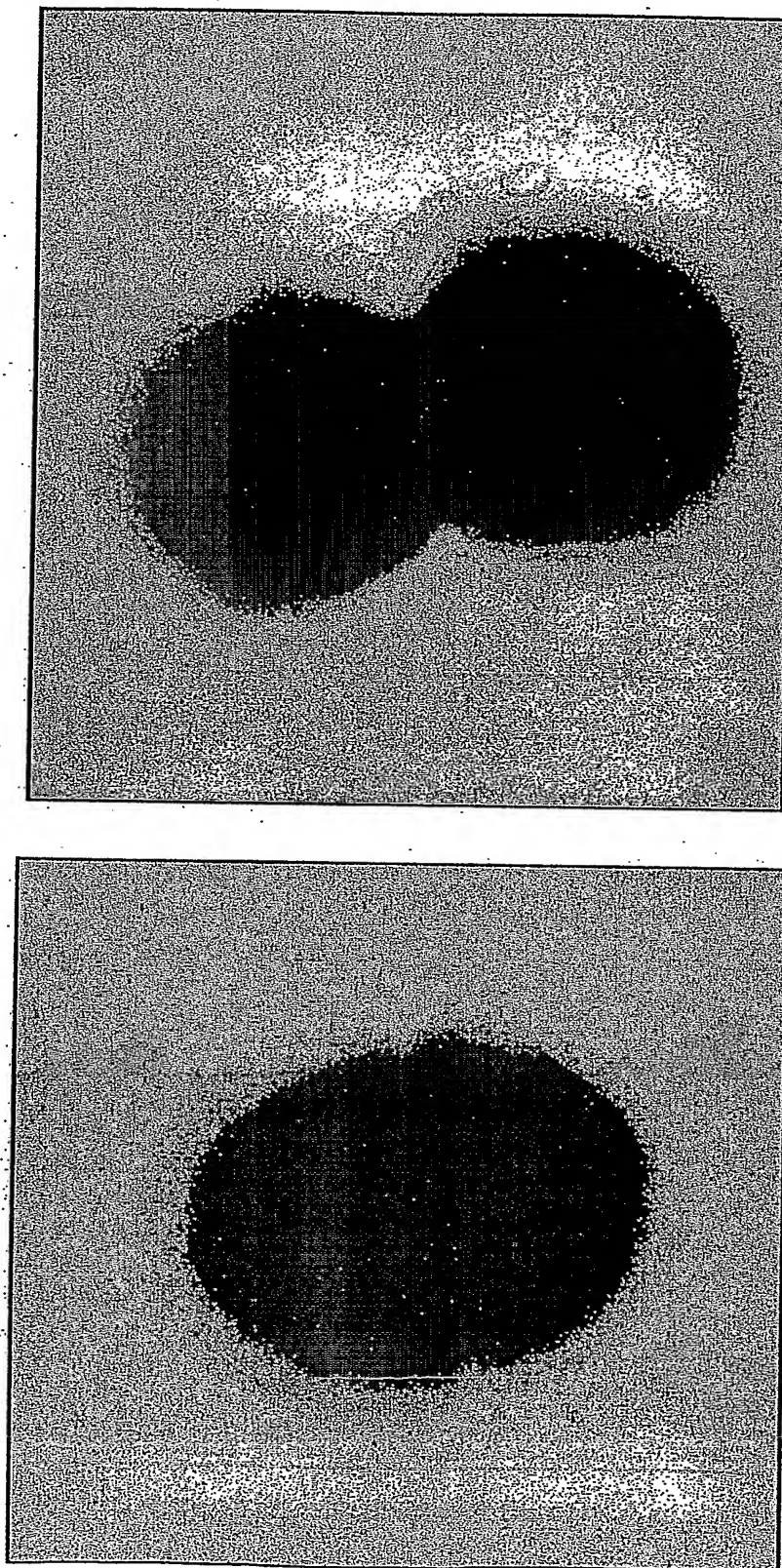


PCT/US05/27239 149/487

# GBS STRAIN COH1 over GBS80

Figure 34

Negative staining EM



## GBS STRAIN COH1 over GBS80

IEM anti-GBS80 (gold particles 10nm)

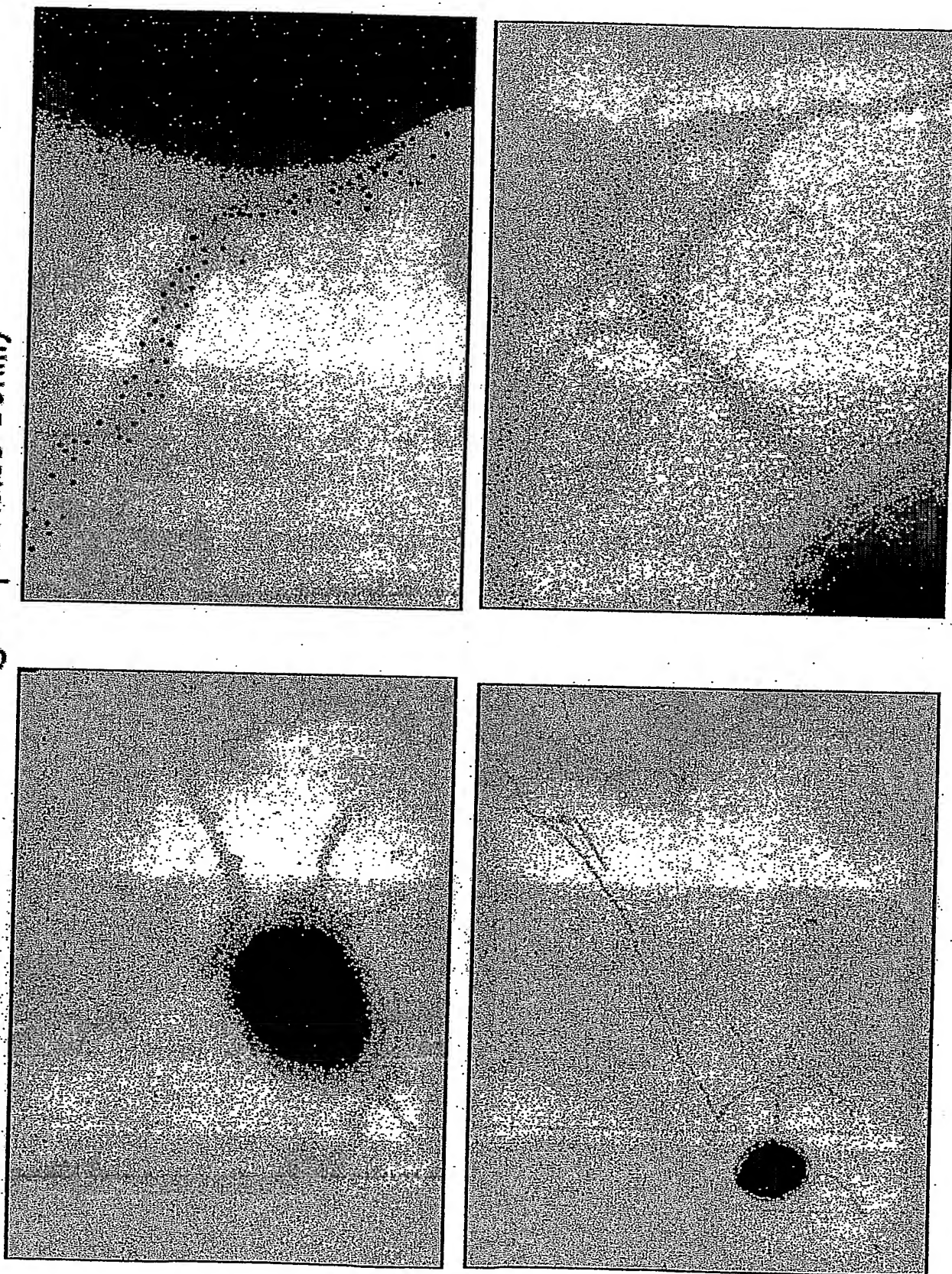


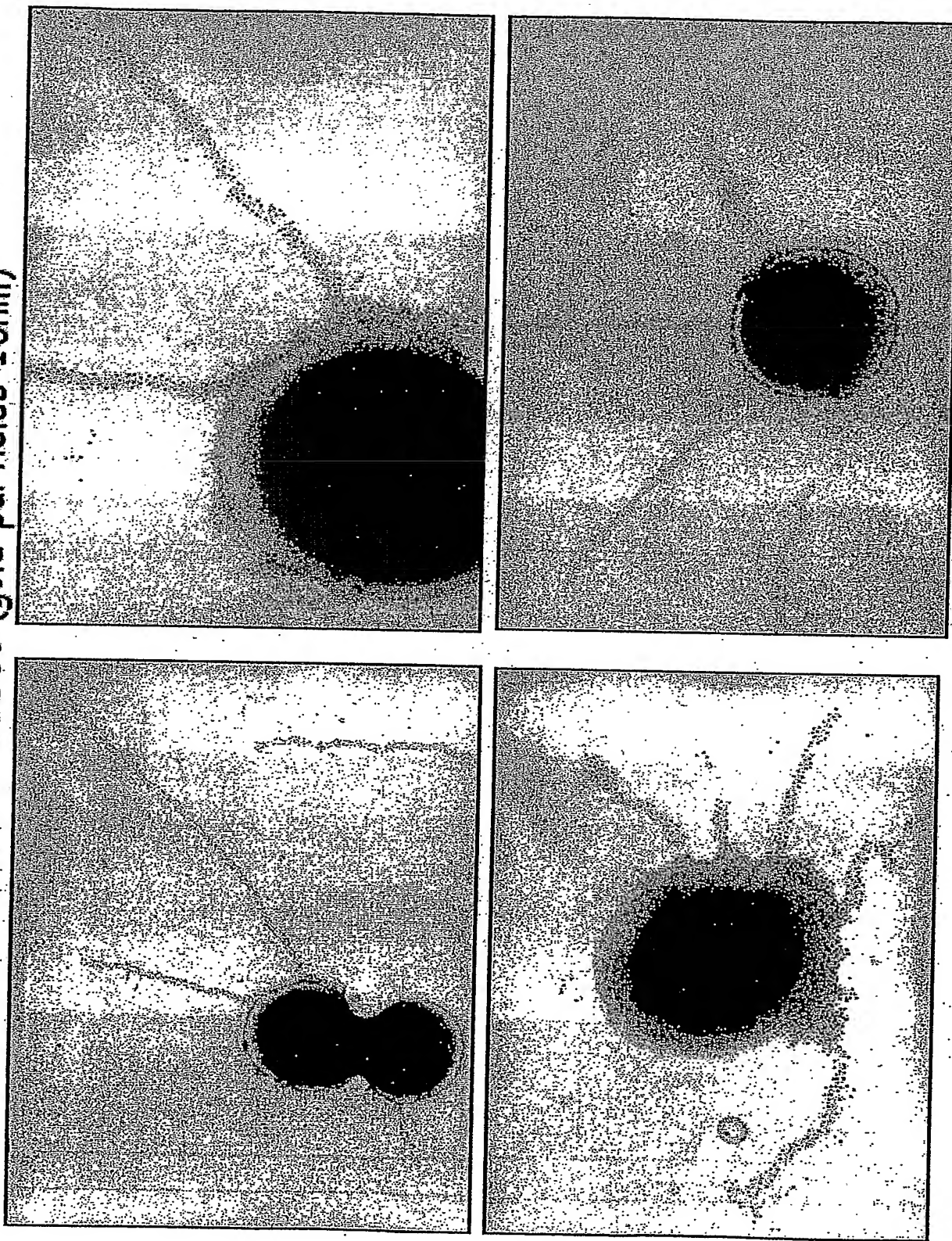
Figure 35



# GBS STRAIN COH1 over GBS80

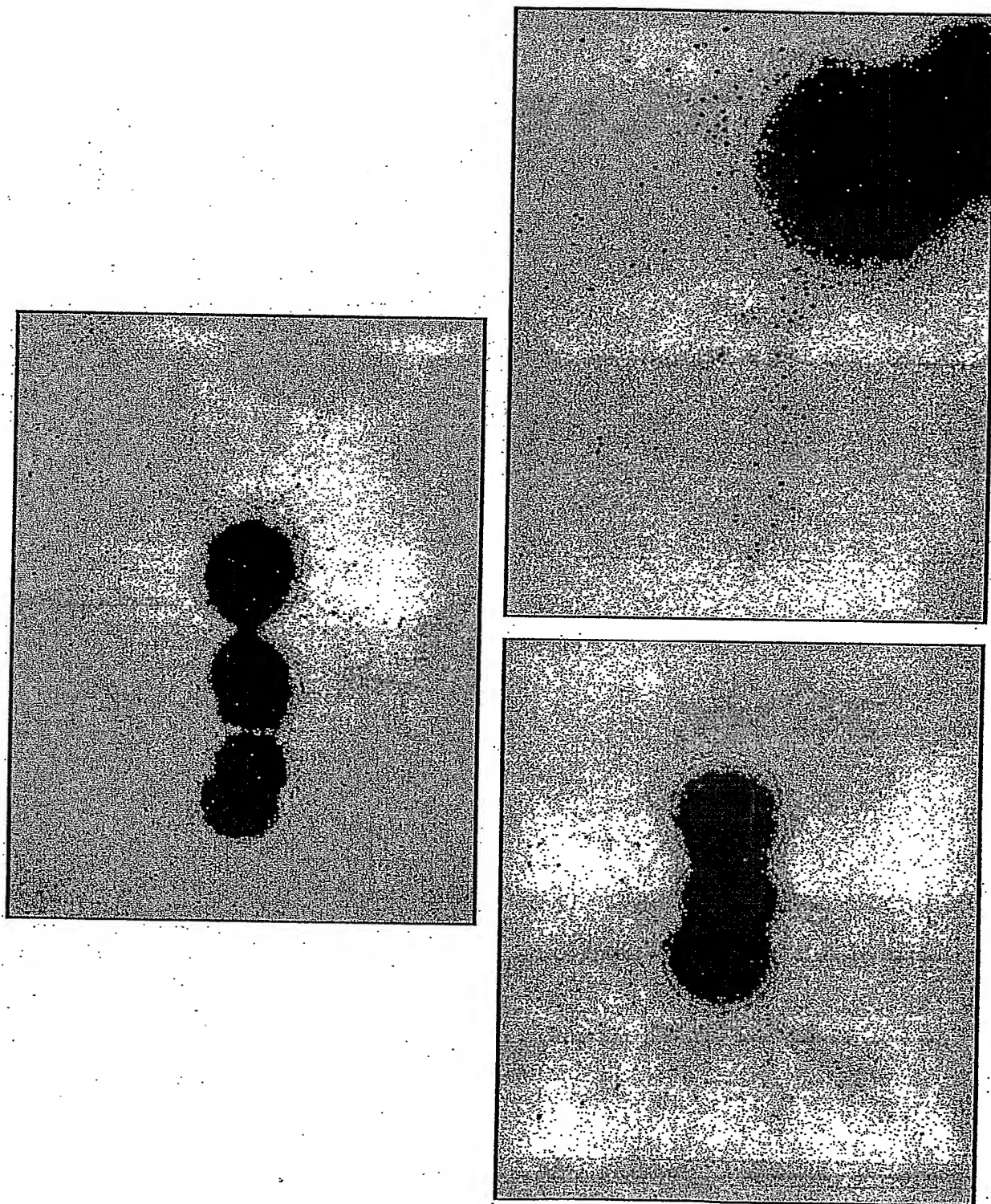
Figure 36

IEM anti-GBS80 (gold particles 10nm)



GBS STRAIN COH1 over GBS80  
IEM anti-GBS80 (gold particles 20nm)

Figure 37





# GBS STRAIN COH1 over GBS80

IEM anti-GBS104 (gold particles 10nm)

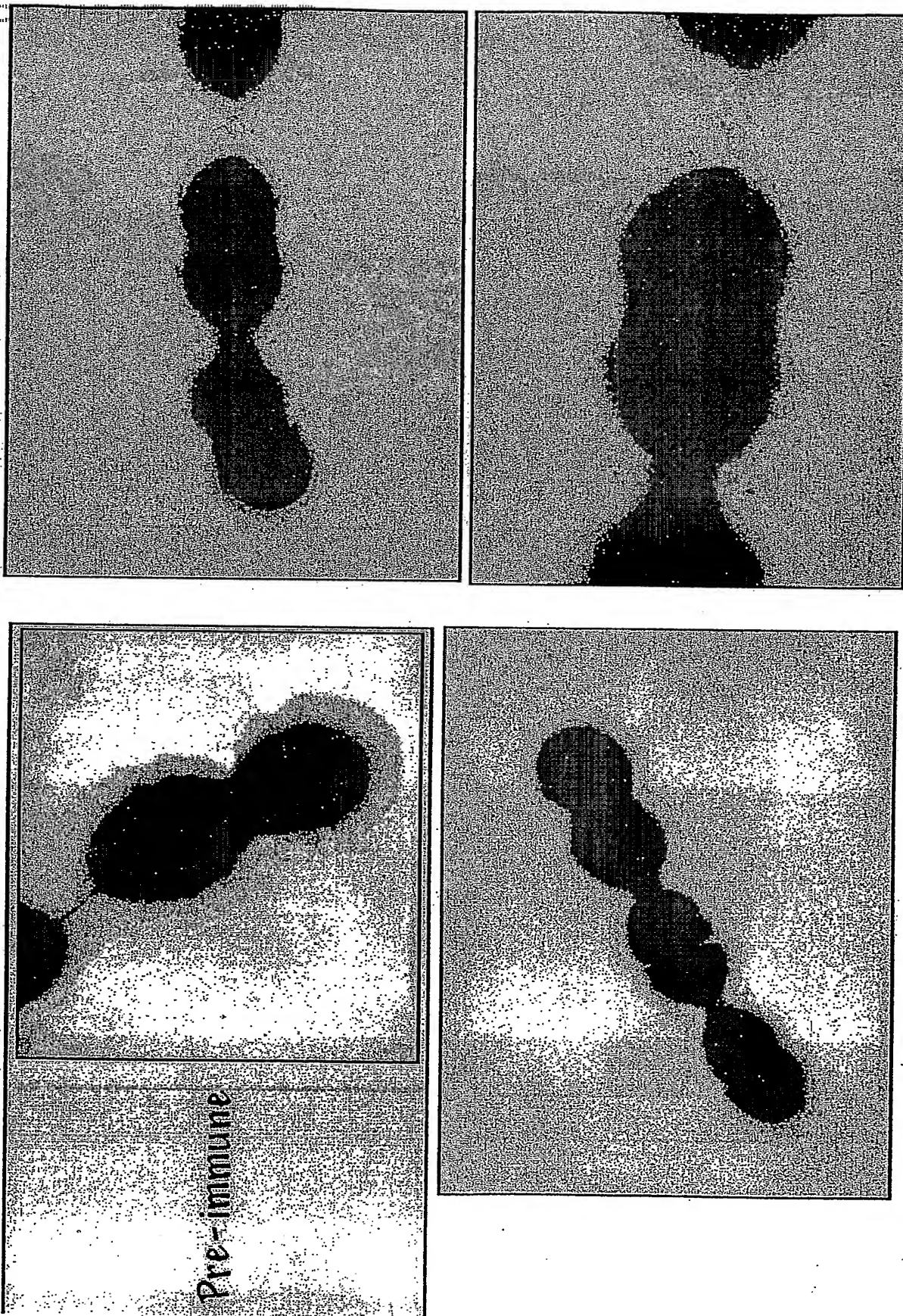
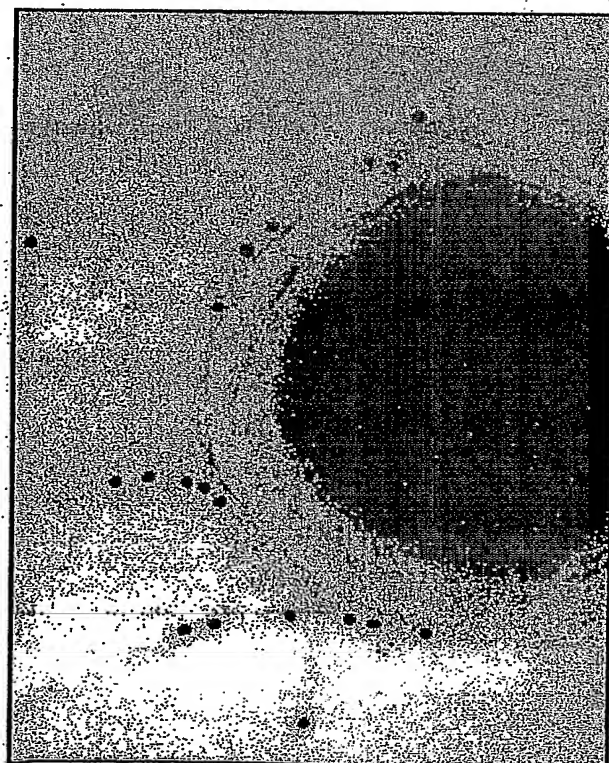
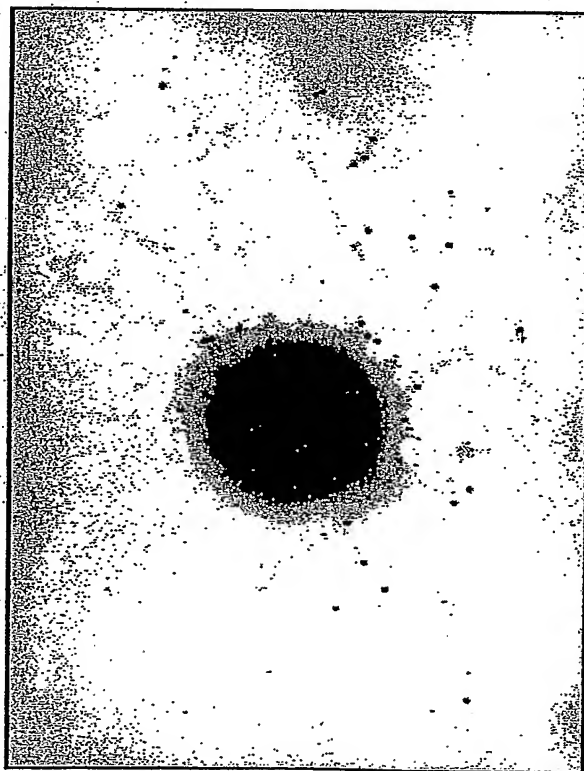
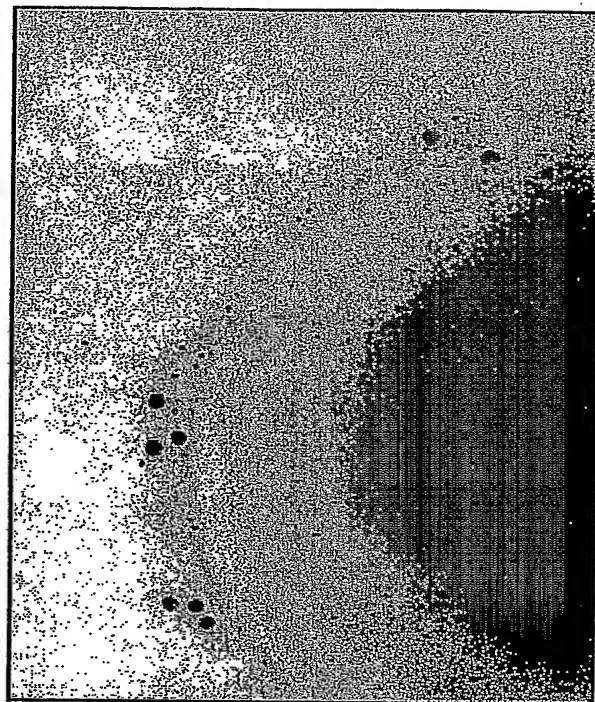
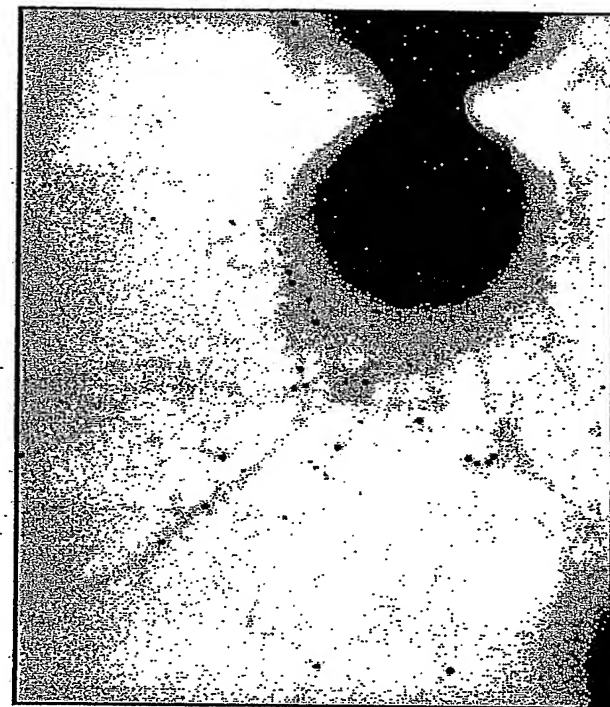


Figure 38

# GBS STRAIN COH1 over GBS80

Figure 39

IEM anti-GBS80 (gold particles 20nm) anti-GBS104 (gold particles 10nm)

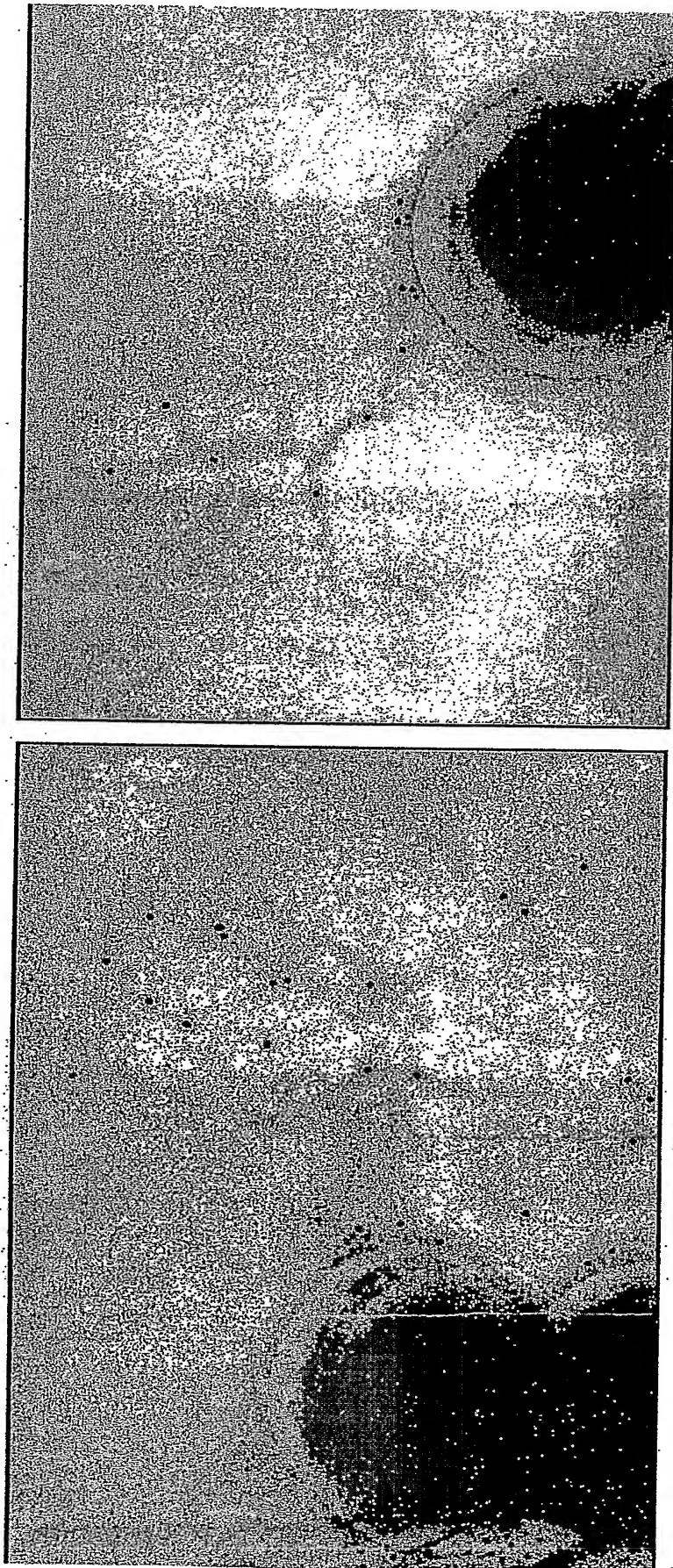




# Figure 40

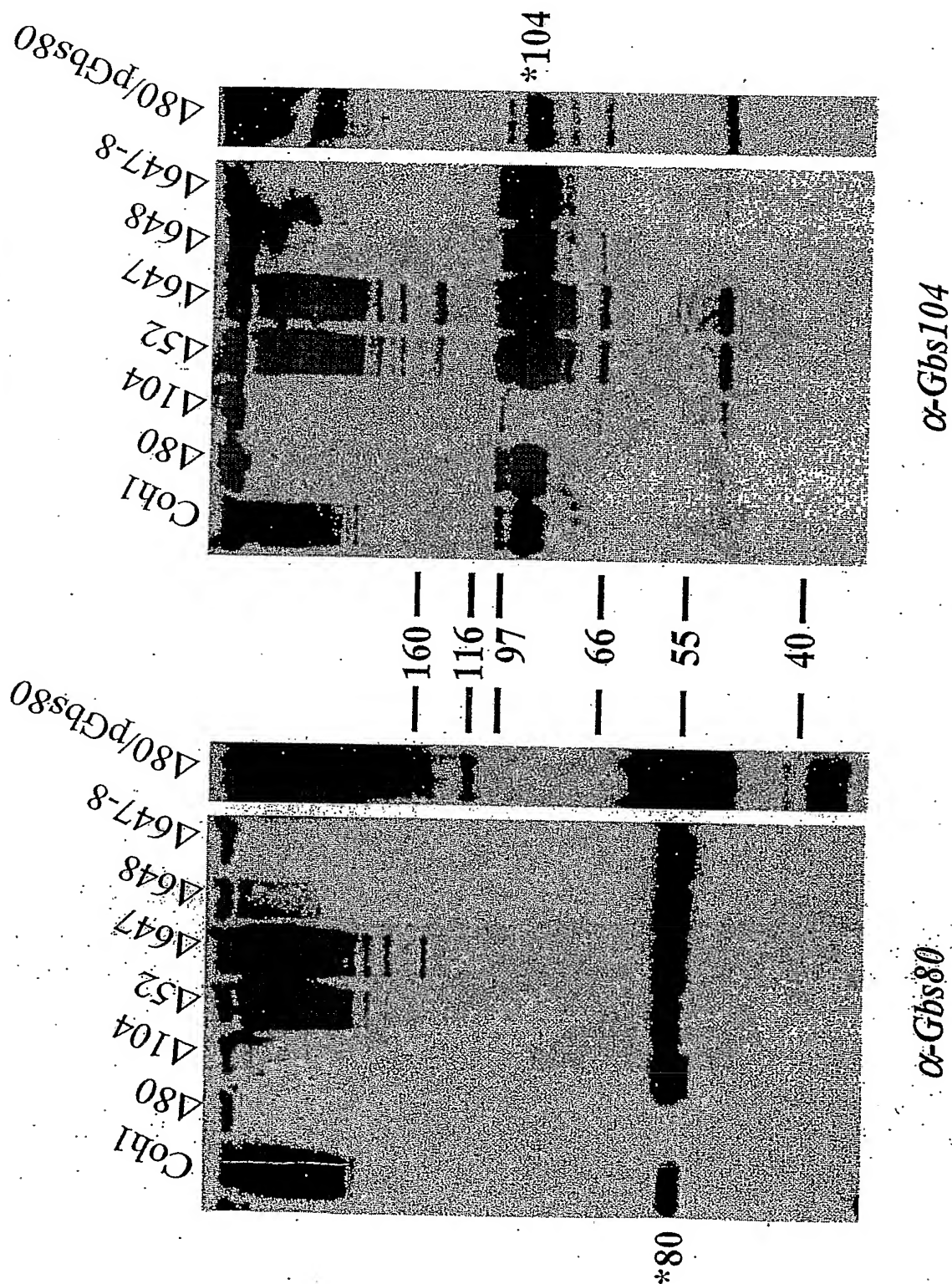
## GBS STRAIN COH1 over GBS80

IEM anti-GBS80 (gold particles 20nm) anti-GBS104 (gold particles 10nm)

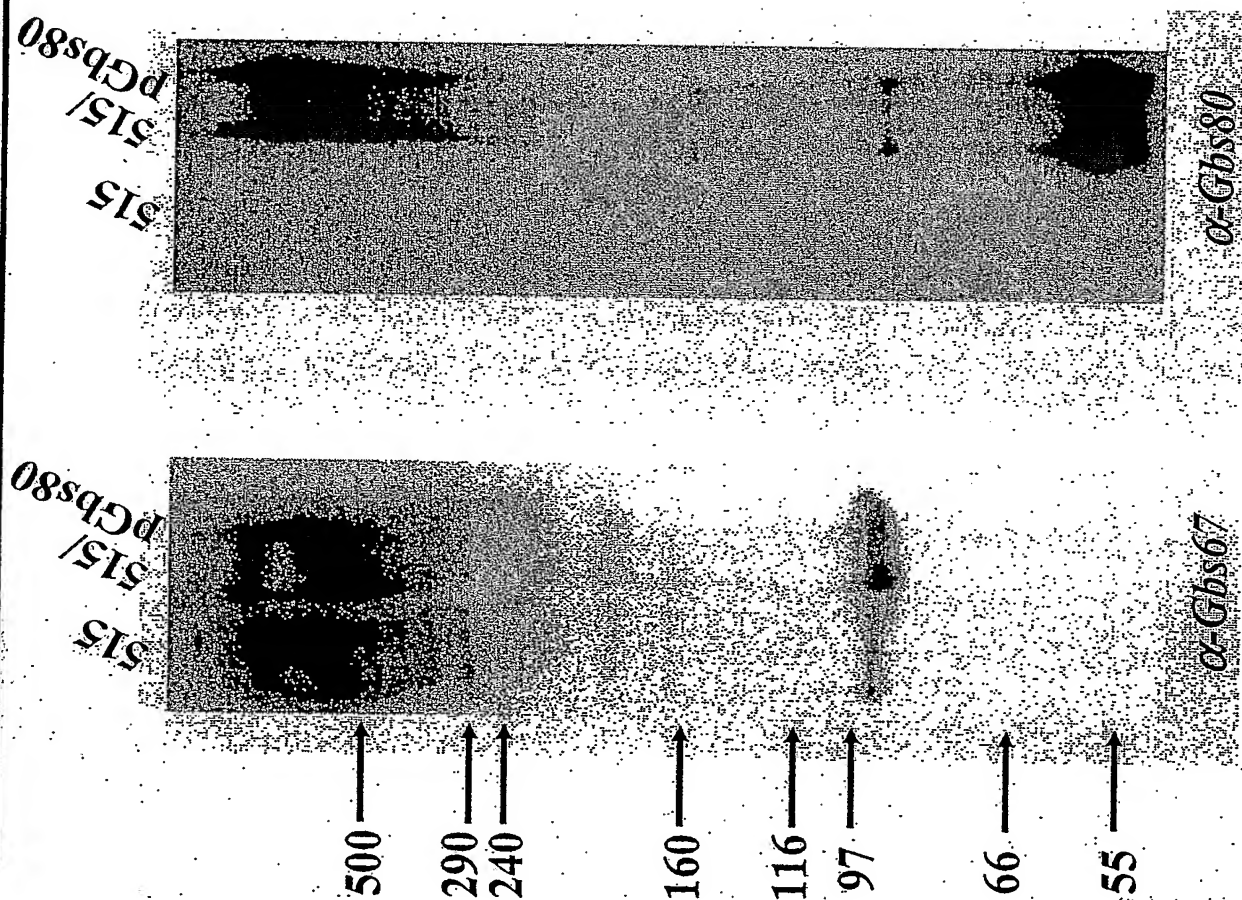


156/487

**Figure 41: GBS 80 is necessary for polymer formation, GBS104 and sortase SAG0648 are necessary for efficient assembly**



PCT/US2005/027239 157/487

**Figure 42: Gbs67 is part of a second pilus;****Gbs80 is polymerized in strain 515****(515 lacks sortase 647-8, but has AI-2 sortases)**



**Figure 43: Two macro-molecules are visible in Coh1 at  $>1000$  kDa, one is the Gbs80 pilin**

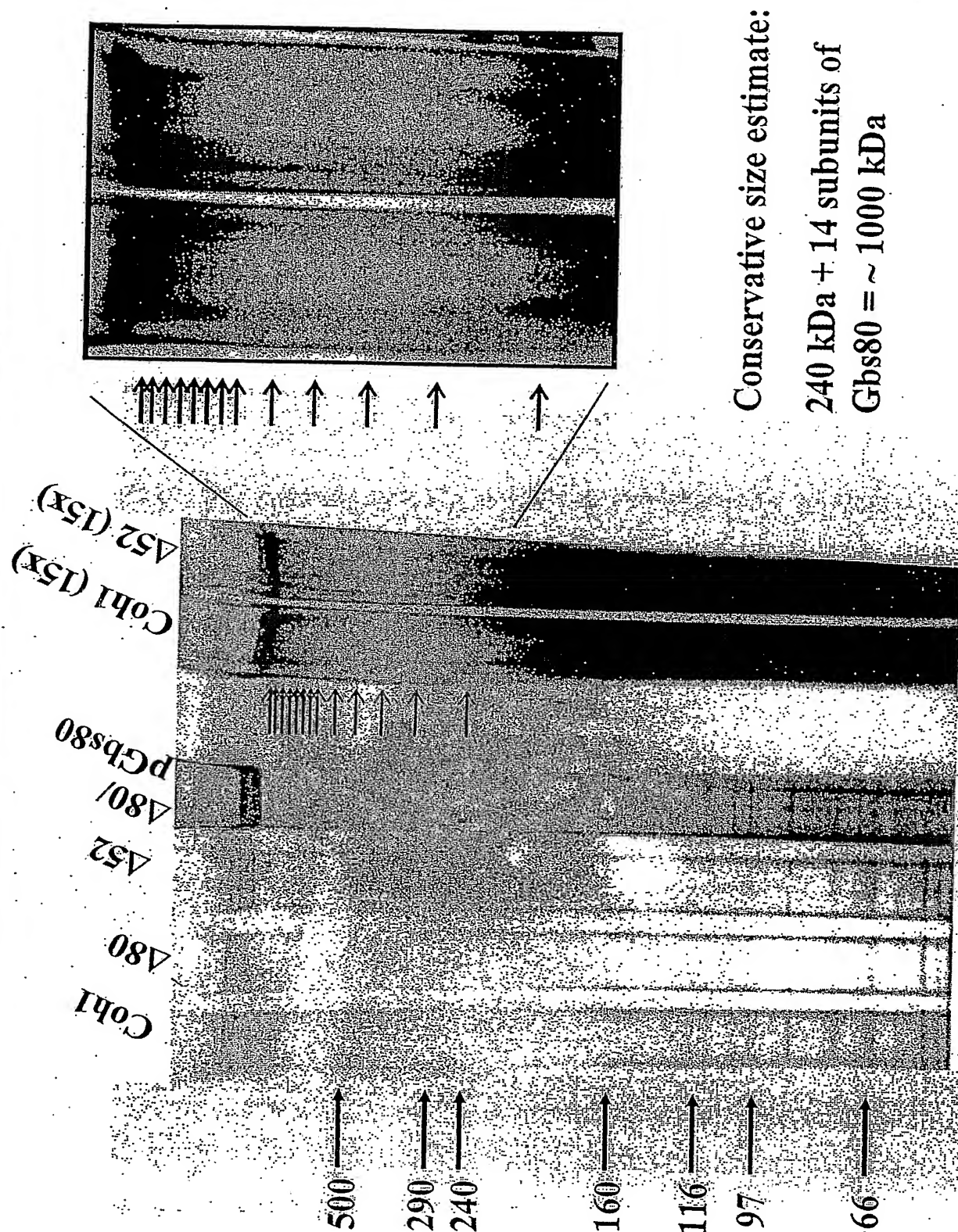


Figure 44

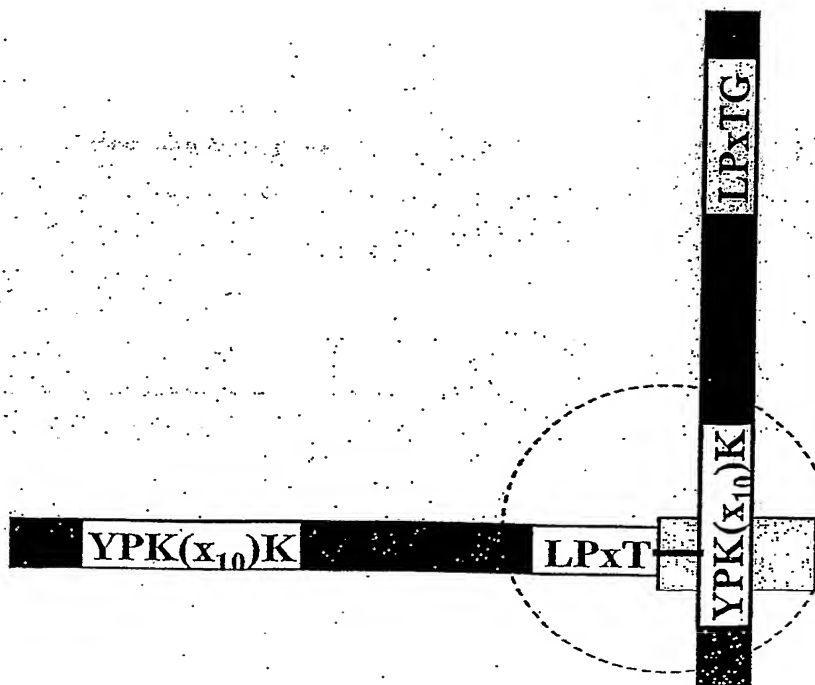
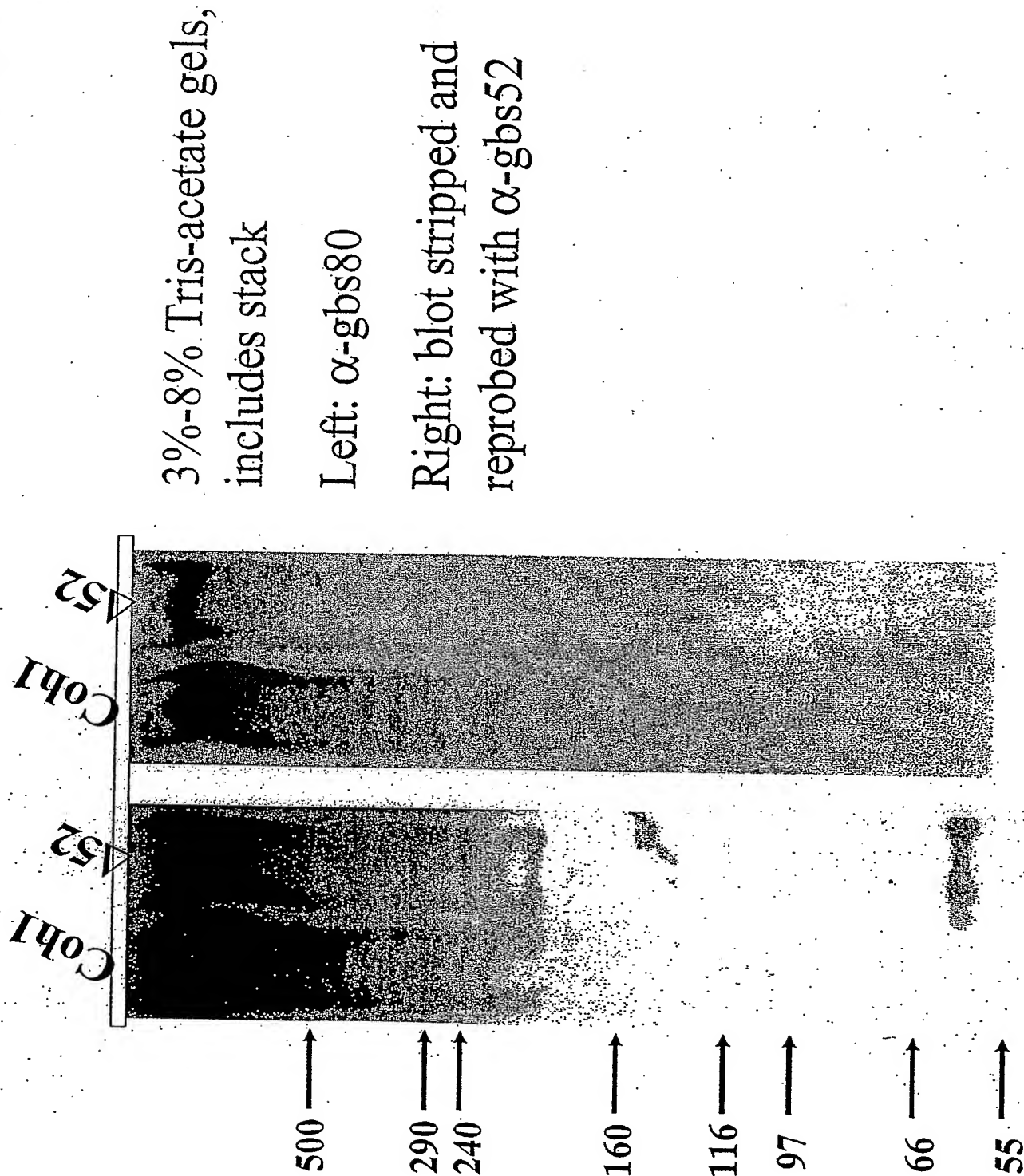
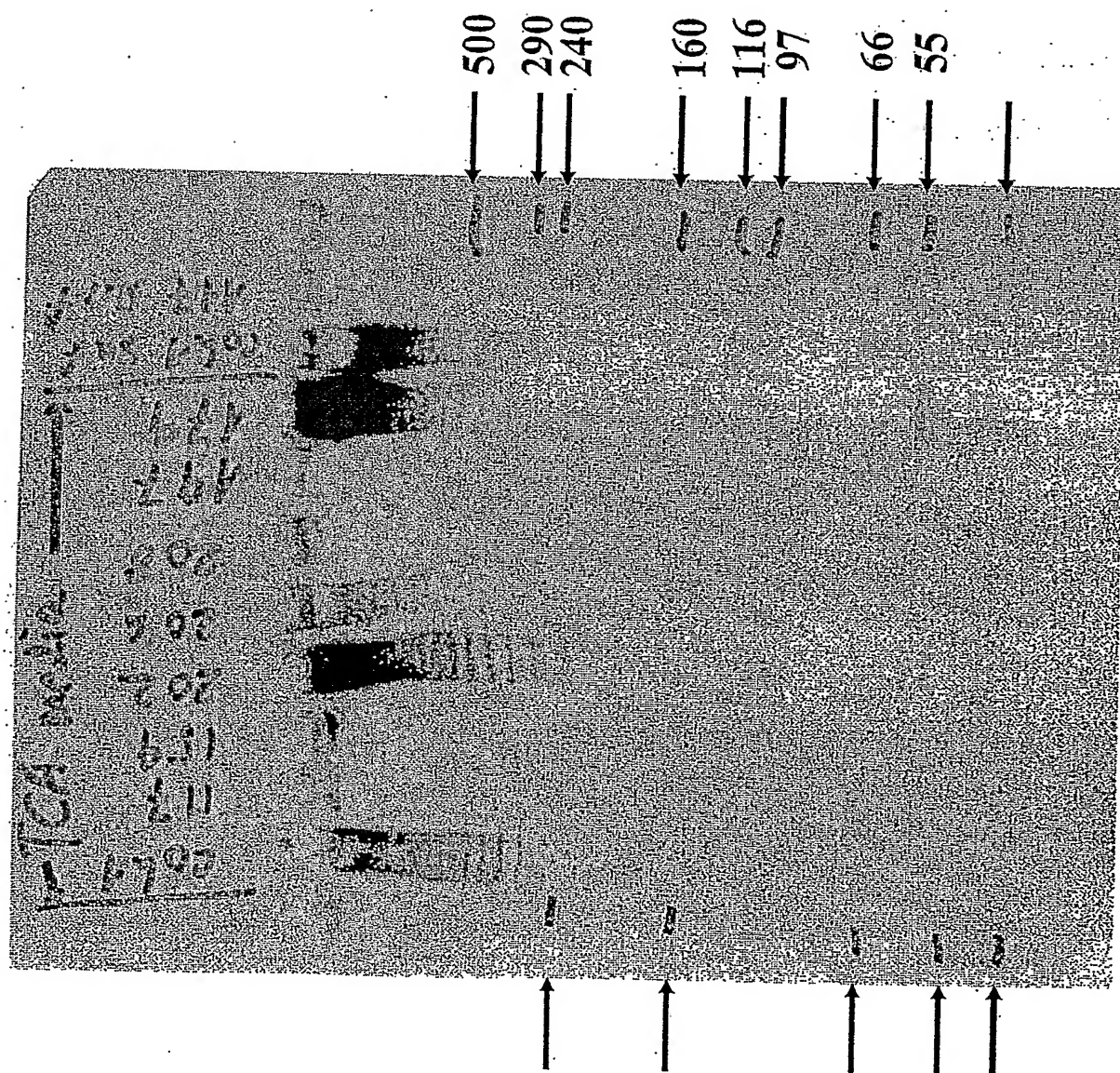


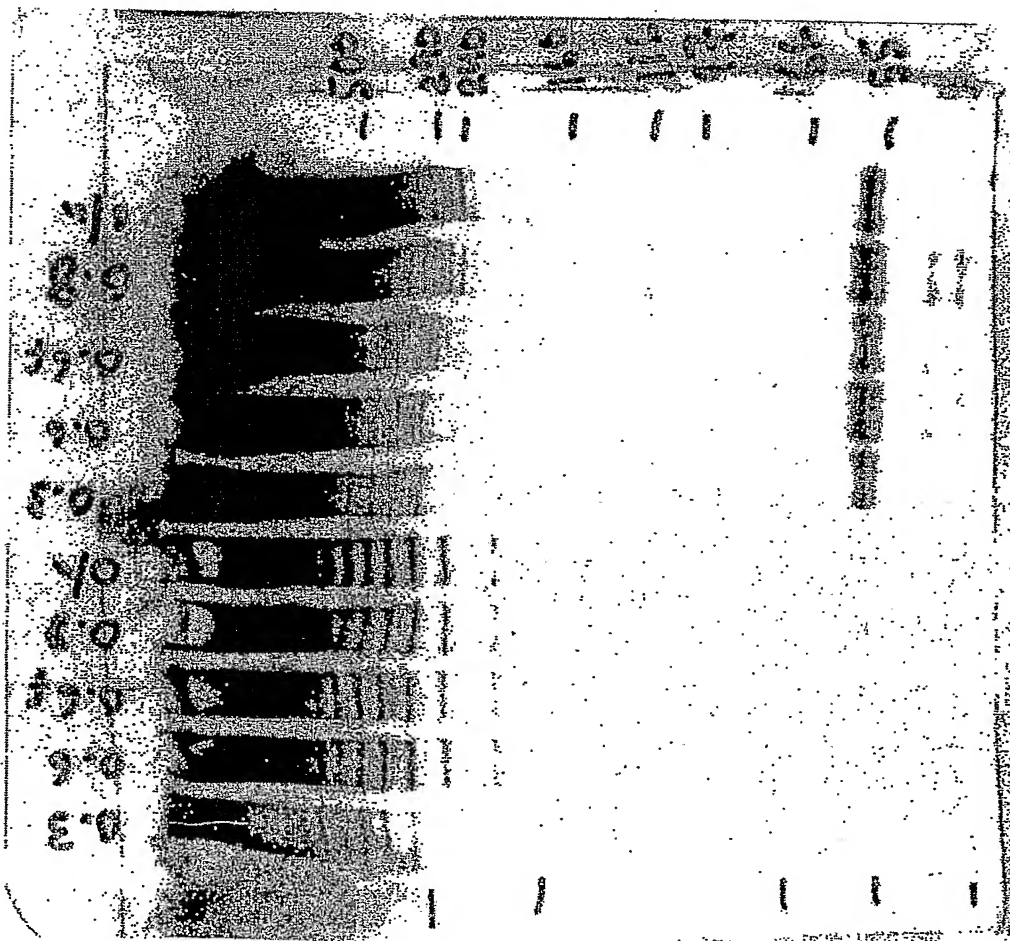
Figure 45: Gbs52 is a minor component of the GBS pilus



**Figure 46:** The pilus is found in the supernatant of the bacterial culture



# Figure 47: The pilus is found in all growth phases of cultures in all growth phases



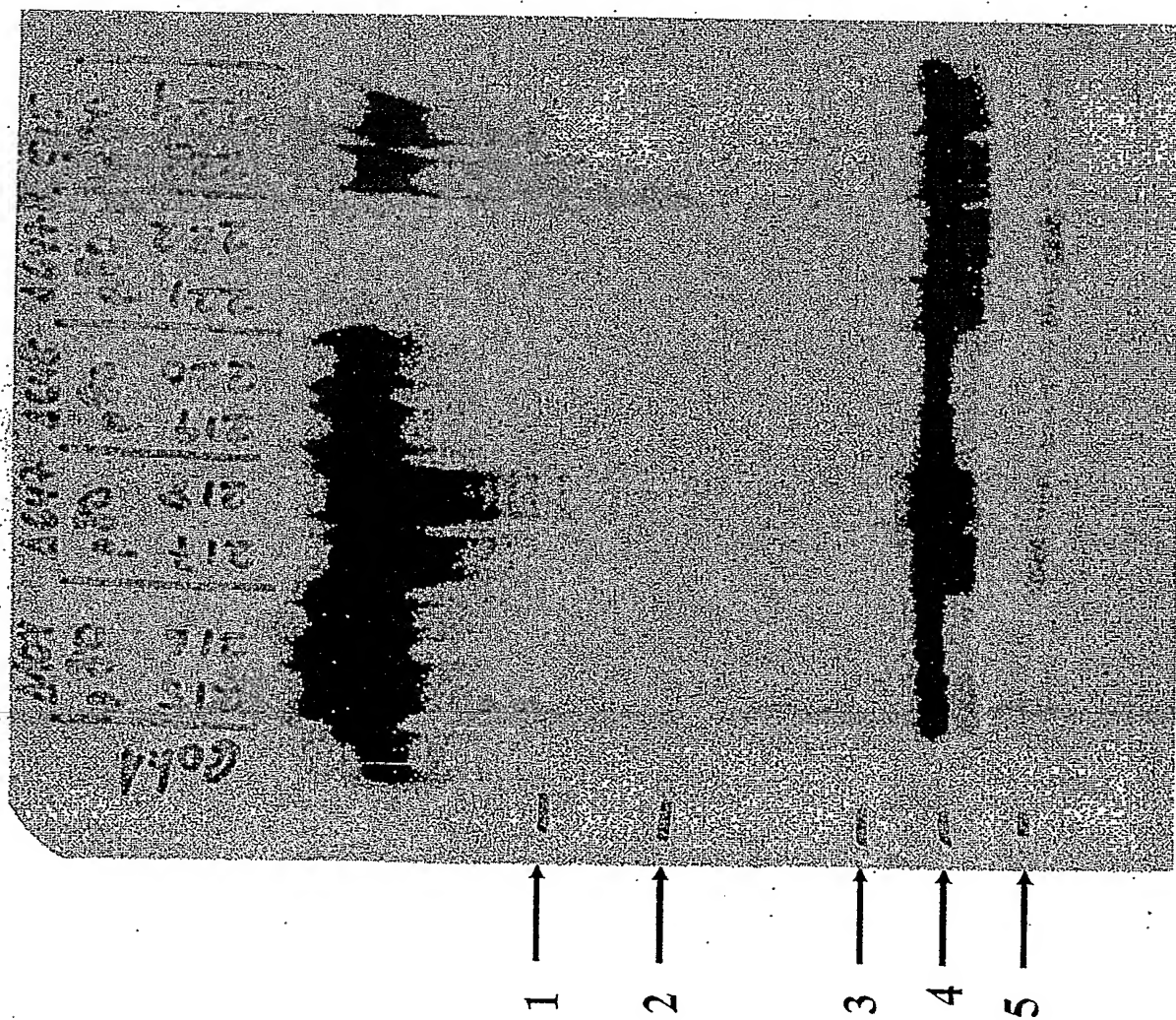
TCA precipitation of 1 ml of THB culture supernatant run on 3-8% SDS-PAGE. OD600 nm are noted above samples, "f" indicates supernatant was filtered (0.2  $\mu$ M syringe filter).

Left five samples: Coh1.

Right five samples: 179 ( $\Delta$ Gbs80/pGbs80).



**Figure 48: In Coh1, only the gbs80 protein and one sortase (sag0647 or sag0648) is required for polymerization**



Over expression of gbs80 in various strain backgrounds (two clones each).

Total protein extract preparations.

Only the double sortase mutant does not polymerize gbs80.

Gbs80 is polymerized in the DK515 strain background (lacks adhesin island 1, adhesin island 2 is 2603-like). Presumably, sag1405&sag1406 are responsible for polymerization.

GBS STRAIN JM9030013

IEM anti-GBS80

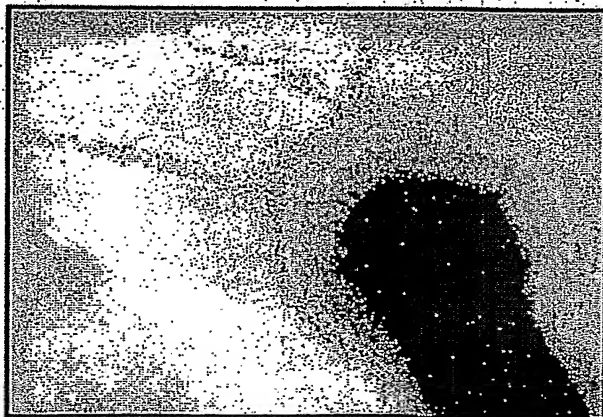
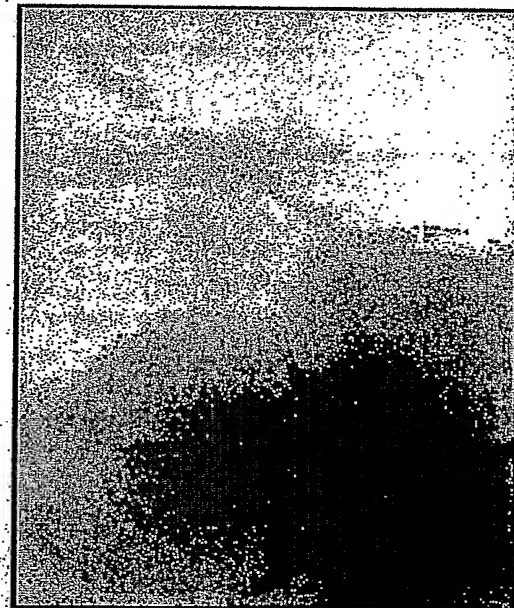
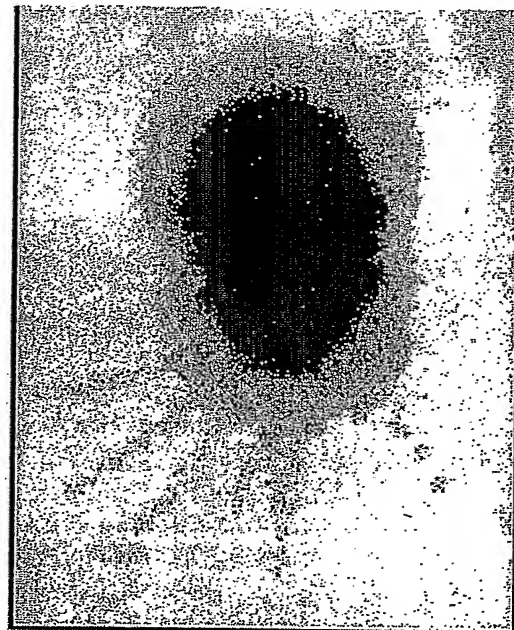
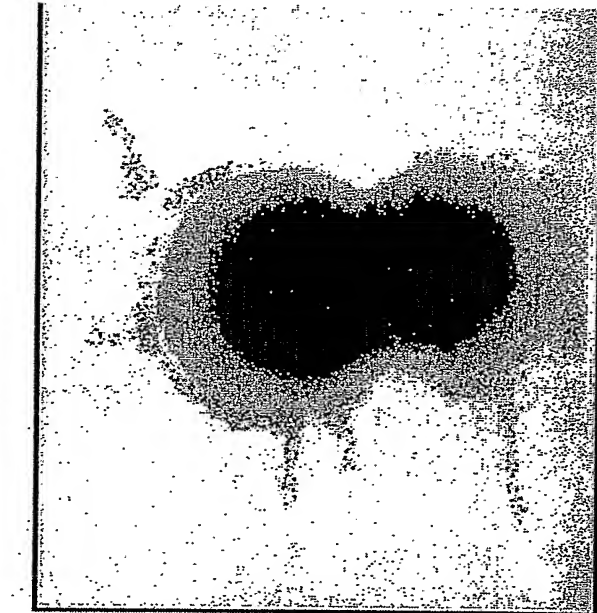
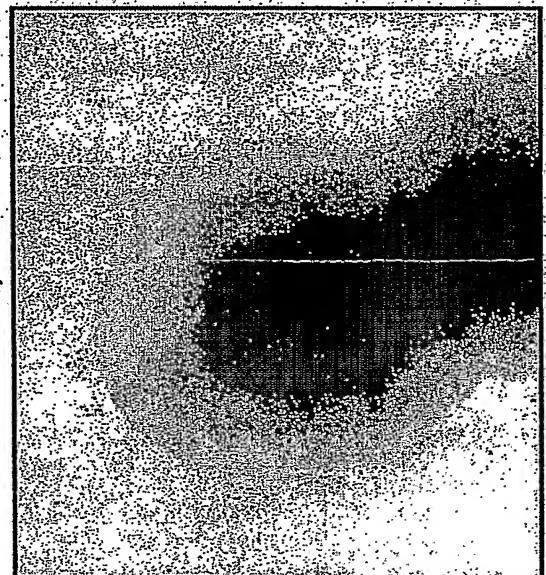
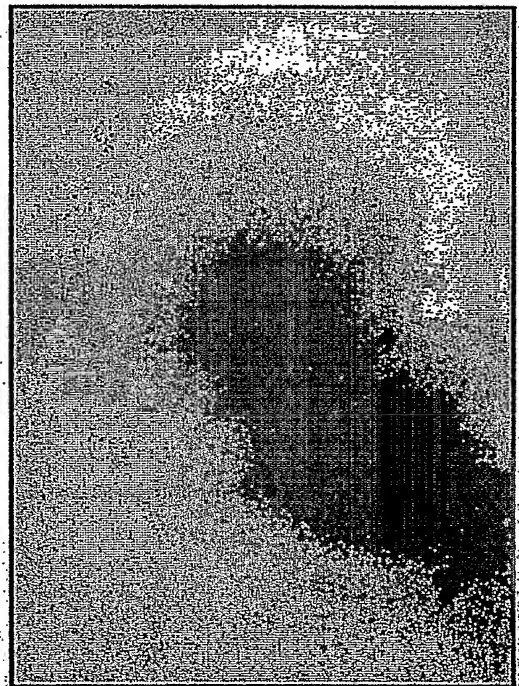
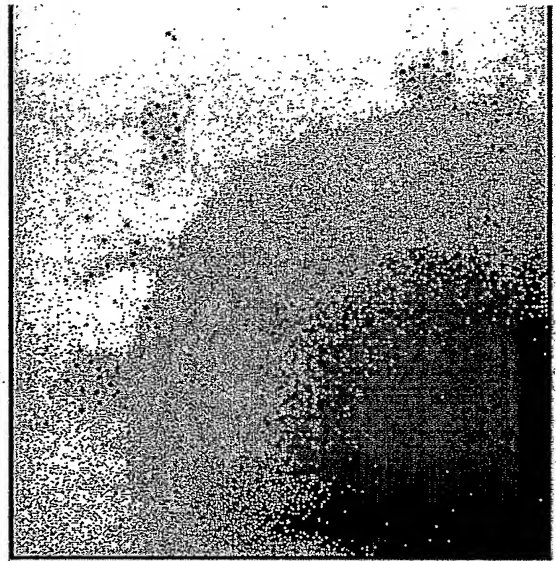
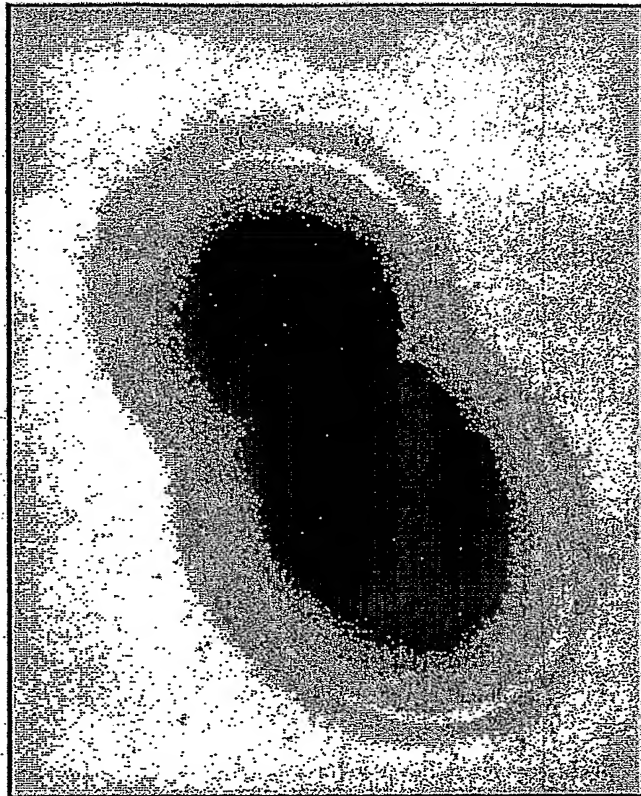


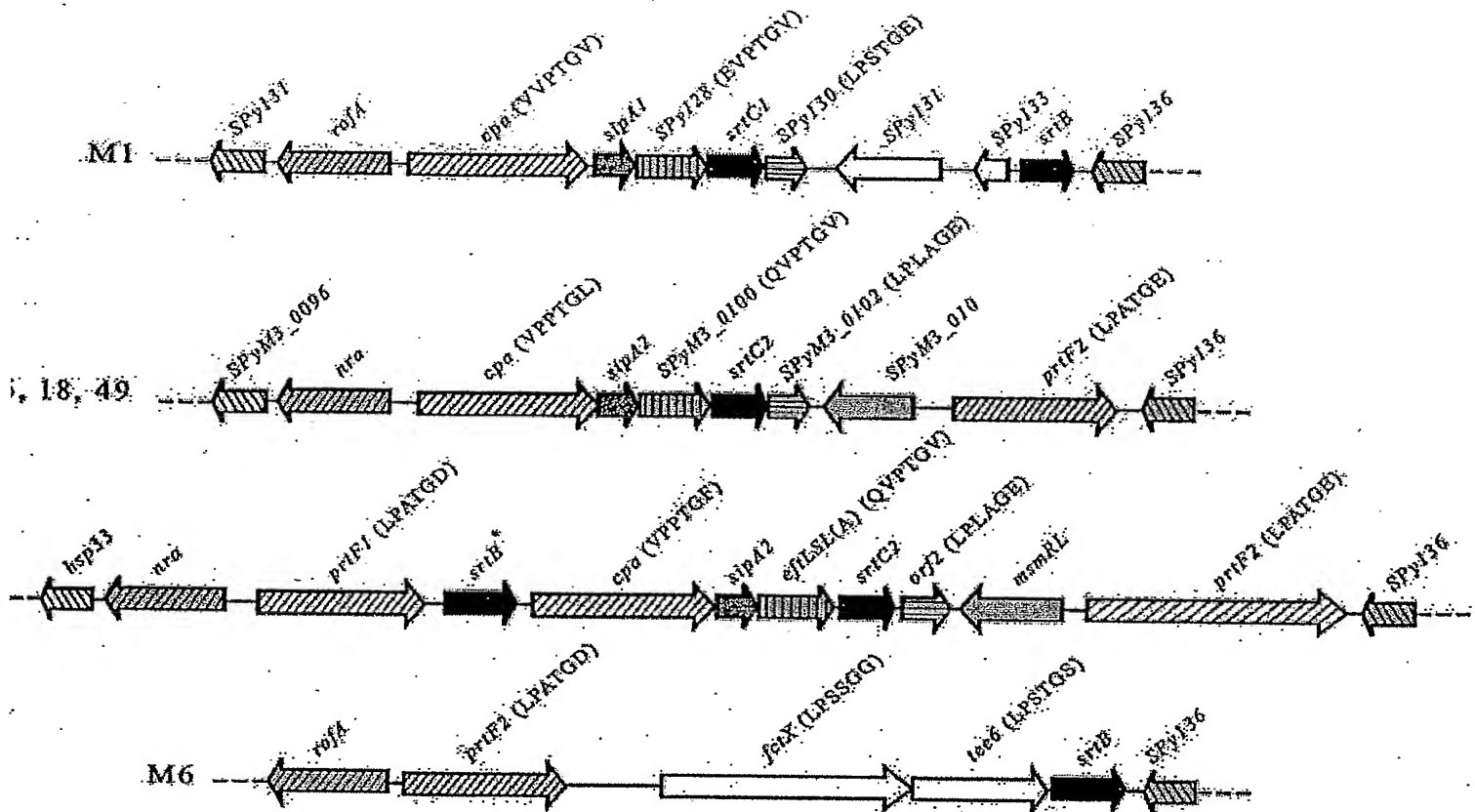
FIGURE 49





GBS STRAIN JM9030013  
IEM anti-GBS104

FIGURE 50



M1	1	MAVRUKKEDKLIIVIVIC-PTIC-LEADL-TH-ALIVYCC-NATINVOOYK-KKGVOFEDLLALNSDVMWILVKSHTIL
M3	1	MAVTIVQVINKAIDTLILIFCLVVLFLAGFGLWDSYHLYCCADASNFRKFKTAQOQPKFEDLLALNEDVIGRLNIPGTHL
M5	1	MAVTIVQVINKAIDTLILIFCLVVLFLAGFGLWDSYHLYCCADASNFRKFKTAQOQPKFEDLLALNEDVIGRLNIPGTHL
M12	1	MAVTIVQVINKAIDTLILIFCLVVLFLAGFGLWDSYHLYCCADASNFRKFKTAQOQPKFEDLLALNEDVIGRLNIPGTHL
M18	1	MAVTIVQVINKAIDTLILIFCLVVLFLAGFGLWDSYHLYCCADASNFRKFKTAQOQPKFEDLLALNEDVIGRLNIPGTHL
M49	1	MAVTIVQVINKAIDTLILIFCLVVLFLAGFGLWDSYHLYCCADASNFRKFKTAQOQPKFEDLLALNEDVIGRLNIPGTHL
M1	80	YPLVQCKTNLEYINKAVEGSSVAMSGSIFLTKNNHNDFTDYSLIYGHMAGNANMFGRIKFKLKKDFFKKNKATLSTKPK
M3	82	YPLVQCKTNLEYINKAVEGSSVAMSGSIFLTKNNHNDFTDYSLIYGHMAGNANMFGRIKFKLKKDFFKKNKATLSTKPK
M5	82	YPLVQCKTNLEYINKAVEGSSVAMSGSIFLTKNNHNDFTDYSLIYGHMAGNANMFGRIKFKLKKDFFKKNKATLSTKPK
M12	82	YPLVQCKTNLEYINKAVEGSSVAMSGSIFLTKNNHNDFTDYSLIYGHMAGNANMFGRIKFKLKKDFFKKNKATLSTKPK
M18	82	YPLVQCKTNLEYINKAVEGSSVAMSGSIFLTKNNHNDFTDYSLIYGHMAGNANMFGRIKFKLKKDFFKKNKATLSTKPK
M49	82	YPLVQCKTNLEYINKAVEGSSVAMSGSIFLTKNNHNDFTDYSLIYGHMAGNANMFGRIKFKLKKDFFKKNKATLSTKPK
M1	161	KLTKIN-PAFCTDAFTSLHFNFLVDYDISSKN-SEHMH-KOKSV-ARELLTINERVALSTCEDDTTIC-RTIVF-GLP
M3	163	KLTVTFACLKTDAPFDQLVFNPNATNQQGKQLVLYISKRSKQKPPV/LKRPKTPVAFSTCENFSTDNKRVIVGCTIQE
M5	163	KLTVTFACLKTDAPFDQLVFNPNATNQQGKQLVLYISKRSKQKPPV/LKRPKTPVAFSTCENFSTDNKRVIVGCTIQE
M12	163	KLTVTFACLKTDAPFDQLVFNPNATNQQGKQLVLYISKRSKQKPPV/LKRPKTPVAFSTCENFSTDNKRVIVGCTIQE
M18	163	KLTVTFACLKTDAPFDQLVFNPNATNQQGKQLVLYISKRSKQKPPV/LKRPKTPVAFSTCENFSTDNKRVIVGCTIQE
M49	163	KLTVTFACLKTDAPFDQLVFNPNATNQQGKQLVLYISKRSKQKPPV/LKRPKTPVAFSTCENFSTDNKRVIVGCTIQE

FIGURE 51

PCT/US05/27239

GI-19224135	1	MNNKKQKQDAPR-VSNRRP	-----KQLTWTLGVFLMFLTLSSMRGAQSIFFGEEK
ORF78	1	QKRDKNYGSANNKR	-----ROTTIGLLKVELTFVALIG-----IVG
GI-21909634	1	QKRDKNYGSANNKR	-----ROTTIGLLKVELTFVALIGIVGFSIRAFG
GI-28810257	1	QKRDKNYGSANNKR	-----ROTTIGLLKVELTFVALIGIVGFSIRAFG
GI-19745301	1	QKRDKNYGSANNKR	-----ROTTIGLLKVELTFVALIGIVGFSIRAFG
GAS15	1	LRGEKMTTRFFNKLNTLTQRVLSKNS	-----RRTFWTLGVFLMFLALVTSVVGAKTVFG
GI-19224135	53	RISEVSVEKIKSPDD--AYPWYGYDSYDSSHFYVERFVAHDLRVNNGSRSYQVYCFNL	
ORF78	39		
GI-21909634	46	AEQSVPNKQSSVQ--DYPWYGYDSYSNGYEDYSPLNTYHNKLVNLDGSREYQAYCFNL	
GI-28810257	46	AEQSVPNKQSSVQ--DYPWYGYDSYSNGYEDYSPLNTYHNKLVNLDGSREYQAYCFNL	
GI-19745301	46	AEQST	
GAS15	58	LVESTENAINPDSSESVRWYGYESYVRGHEFYKQFVAHDLRVNLEGSRSYQVYCFNL	
GI-19224135	111	NSHPEKQKQFSKQENRVDGTGEVFTNYEQTPKIRGESLNNKLLSIMYNAFKNANGVM	
ORF78	39		
GI-21909634	103	TKHFFSKSDSVRSQWYKRLGCTNENFIKLADKPRIEDGQLOONILRTIYNGYPNDRNGIM	
GI-28810257	103	TKHFFSKSDSVRSQWYKRLGCTNENFIKLADKPRIEDGQLOONILRTIYNGYPNDRNGIM	
GI-19745301	52		
GAS15	117	KNAFFLGSDSVKKYKREHDIKSTKEEDYMSFRTGTGDELNOKTRAVMYNGHQNANGIM	
GI-19224135	171	DKLEPLNAILVTOQAWWYSDSSYGN-IKTLWASELKDGKIDFEQKLMREAYSRIISDD	
ORF78	39	-----ESIRAG	
GI-21909634	163	KGIDPLNAILVTQNAIYVTDSSYISDTSKAFQCEETDLKIDSQOLMLRNALKRLINEK	
GI-28810257	163	KGIDPLNAILVTQNAIYVTDSSYISDTSKAFQCEETDLKIDSQOLMLRNALKRLINEK	
GI-19745301	52		
GAS15	177	EGLEPLNAIRVTQBAWYSDNAPLISNPDSESKRESESNLYSTISQLSLNRQALKRLIDEN	
GI-19224135	230	LEETSKNKLFGQSKLNIIVPQDNS-----VONLLSAEYVPESPEAFQOSIEPEVOTKKT	
ORF78	46		-----AEEKSTETKKT
GI-21909634	223	EVESLPNQVPANYSLSIFQSSDNT-----FQNLLEAEYVPDTPPKPG	
GI-28810257	223	EVESLPNQVPANYSLSIFQSSDNT-----FQNLLEAEYVPDTPPKPG	
GI-19745301	52		-----ETKKT
GAS15	237	LATKMPKQVPDBEQLSIFESDNGDKYNGKYQNLLESGGVETKPEETPGDPPMPPNQOTT	
GI-19224135	284	SVIIRKYAEGDYSKLEGGATLRLTGEDILDFQEKVFSNGTGEKIELSNGTYTLTETSSP	
ORF78	57	SVIIRKYAEGDYSKLEGGATLRLTGEDILDFQEKVFSNGTGEKIELSNGTYTLTETSSP	
GI-21909634	275	SVIIRKYAEGDYSKLEGGATLRLTGEDILDFQEKVFSNGTGEKIELSNGTYTLTETSSP	
GI-28810257	275	SVIIRKYAEGDYSKLEGGATLRLTGEDILDFQEKVFSNGTGEKIELSNGTYTLTETSSP	
GI-19745301	57	SVIIRKYAEGDYSKLEGGATLRLTGEDILDFQEKVFSNGTGEKIELSNGTYTLTETSSP	
GAS15	297	SVIIRKYAEGDYSKLEGGATLRLTGEDILDFQEKVFSNGTGEKIELSNGTYTLTETSSP	
GI-19224135	344	DGYKIAEPIKFRVNVKRVFIQKDGSOVENPNKEVAEPYSVBYSDQDSNYINPETETP	
ORF78	117	DGYKIAEPIKFRVNVKRVFIQKDGSOVENPNKEVAEPYSVBYSDQDSNYINPETETP	
GI-21909634	335	DGYKIAEPIKFRVNVKRVFIQKDGSOVENPNKEVAEPYSVBYSDQDSNYINPETETP	
GI-28810257	335	DGYKIAEPIKFRVNVKRVFIQKDGSOVENPNKEVAEPYSVBYSDQDSNYINPETETP	
GI-19745301	117	DGYKIAEPIKFRVNVKRVFIQKDGSOVENPNKEVAEPYSVBYSDQDSNYINPETETP	
GAS15	357	AGYSIAEPIKFRVNVKRVFIQKDGSOVENPNKEVAEPYSVBYSDQDSNYINPETETP	
GI-19224135	404	YGRFYAKNKGSSQVYCFNADLHSPFSEDGGGTIDPLISTMKEVKYTHIAGSDFRY	
ORF78	174	YAKFYGRNYDGSQIVCFNANLKSPPDSEDHGATINPDFTTGDIRYSHIAGSDFRY	
GI-21909634	393	YGRFYAKNKGSSQVYCFNADLHSPFSEDGGGTIDPLISTMKEVKYTHIAGSDFRY	
GI-28810257	393	YGRFYAKNKGSSQVYCFNADLHSPFSEDGGGTIDPLISTMKEVKYTHIAGSDFRY	
GI-19745301	177	YGRFYAKNKGSSQVYCFNADLHSPFSEDGGGTIDPLISTMKEVKYTHIAGSDFRY	
GAS15	413	YAKFYAKNKGSSQVYCFNADLHSPFSEDGGGTIDPLISTMKEVKYTHIAGSDFRY	

FIGURE 52

PCT/US05/27239

GI-19224135 464 ALRPRDTNPEDFLKHINKVTEKGYNKKGD--SYNGLTEIOPRAATQLAIYYFTDSIDLRT  
ORF78 233 ANTARDEDFQLFKHVNIENGYHKKGQAIPYNSLTBAQFRAATQLAIYYFTDSVDL--  
GI-21909634 453 AATPRDKDADFFLKHINKILDKGYKKKGD--TYKILTEAQFRAATQLAIYYFTDSADLIT  
GI-28810257 453 AATPRDKDADFFLKHINKILDKGYKKKGD--TYKILTEAQFRAATQLAIYYFTDSADLIT  
GI-19745301 237 ANNPRASINDELSQVNIIVLEKGYRDDST--TYANLPSVEFRAATQLAIYYFTDSVDLON  
GAS15 472 TVKPRDTDPDITFLKHINKVIEKGYREKQAIPYNSLTETQLRAATQLAIYYFTDSABL--

GI-19224135 522 LKTYNNGKGYHGFESWDEKTLAVTKELINYAOD--NSAPQLTNLDFFVPNNISKYQSLIGTE  
ORF78 291 --TRDRLNDEHGFCDMDNDOTLGVAKKIIVEYALS--DEDSKLTNLDFFVPNNISKYQSLIGTE  
GI-21909634 511 LKTYNDNKGYHGFCDKLDATLAVVHELTITYAED--VILEMTQNLDFFVPNNISKYQSLIGTO  
GI-28810257 511 LKTYNDNKGYHGFCDKLDATLAVVHELTITYAED--VILEMTQNLDFFVPNNISKYQSLIGTO  
GI-19745301 295 LADY-----HGFALTTEALNATILETYAEDRANLPNISNLDFFVPNNISKYQSLIGTO  
GAS15 531 ---KDKLNDYHGFCDMDNDSTLAVANILMEYAOD--SNPQLTDLDFPIPNNNISKYQSLIGTO

GI-19224135 581 YHPDDLVDVIRMEDNKKQEVIPVTHSLTVKRTVVGRLGDNTNGFQFELBLNDKTSQPIVNT  
ORF78 348 YHPDDLVDVIRMEDNKKQEVIPVTHSLTVKRTVVGRLGDNTNGFQFELBLNDKTSQPIVNT  
GI-21909634 570 YHPNELIDVISMEDNKKQPIIPITHKLTISKTVGTIADNKKKEFNFEIHLKSSDQQAISGT  
GI-28810257 570 YHPNELIDVISMEDNKKQPIIPITHKLTISKTVGTIADNKKKEFNFEIHLKSSDQQAISGT  
GI-19745301 349 YHPESLVDIIRMEDNKKQPIIPITHKLTISKTVGTIADNKKKEFNFEIHLKSSDQQAISGT  
GAS15 587 YHPEDLVDIIRMEDNKK--EVIPVTHSLTVKRTVVGRLGDNTNGFQFELBLNNKQELISOT

GI-19224135 641 LKTNNOQLVAKDCKYSFNLRHGDITRIEGLFTGYSYTLKETE--KDYIVTVVDNNVSOEFAQS  
ORF78 408 LKTNNOQLVAKDCKYSFNLRHGDITRIEGLFTGYSYTLKETE--KDYIVTVVDNNVSOEFAQS  
GI-21909634 630 YFTNSGELTVTDGKATFLLDGSGLIVEGLPSGYSYEITETGASDYEVSVNGKNAPDGKA  
GI-28810257 630 YFTNSGELTVTDGKATFLLDGSGLIVEGLPSGYSYEITETGASDYEVSVNGKNAPDGKA  
GI-19745301 409 YFTNSGELTVTDGKATFLLDGSGLIVEGLPSGYSYEITETGASDYEVSVNGKNAPDGKA  
GAS15 646 VKTIDKTNLEFFKDGKATINLRHGSLLHLOGLEFGYSYLVKETE--SEGKVKVNSQEVANATV

GI-19224135 701 ASENVTADKEVTFENRKDLVPPTGLTTDGAIVLWLLLLVFFGLLWVLFGRKGLKND--  
ORF78 468 ASENVTADKEVTFENRKDLVPPTGLTTDGAIVLWLLLLVFFGLLWVLFGRKGLKND--  
GI-21909634 690 TKASVKEDETVAFENRKDLVPPTGLTTDGAIVLWLLLLVFFGLLWVLFGRKGLKND--  
GI-28810257 690 TKASVKEDETVAFENRKDLVPPTGLTTDGAIVLWLLLLVFFGLLWVLFGRKGLKND--  
GI-19745301 469 TKASVKEDETVAFENRKDLVPPTGLTTDGAIVLWLLLLVFFGLLWVLFGRKGLKND--  
GAS15 706 SKTGITTSDETVAFENRKDLVPPTGLTTDGAIVLWLLLLVFFGLLWVLFGRKGLKND--

FIGURE 52A



169/487

GI-19224134 WO 2006/078318 PCT/US2005/027239

GI-50913503 SYMFARGEKMNNTPLNKEAGFLVHTKRRKRFVTLVGVFFLLACAGAIGFGQVAY  
MVSSYMFVARGEKMNNTPLNKEAGFLVHTKRRKRFVTLVGVFFLLACAGAIGFGQVAY

GI-19224134 61 AADERTVFNFKSPDDYPPIYGYDSV-----RCIEARYHNLYVNLKGSREYQAYCFNITK  
GI-50913503 61 AADERTVFNFKSPDDYPPIYGYDSV-----RCIEARYHNLYVNLKGSREYQAYCFNITK

GI-19224134 115 YFPRPTYSTTNNEKKIDGSGSAFKSYAANPRVLDENLDKLEKNILNVTYNGVKSNAAGF  
GI-50913503 121 NNPQKNSFIKNNKKIEGCKSFVDYAHTTKEGKE---ELEQRELSELEYNEYPNDANGY

GI-19224134 175 MNGLEDLNAITVTQYAVIHYSDNS-QYQFETLHESEANEKISRQVTLMLREALKKLIDP  
GI-50913503 178 MNGLEDLNAITVTQYAVIHYSDNS-QYQFETLHESEANEKISRQVTLMLREALKKLIDP

GI-19224134 235 NLEATAANKIPSGYRLNIFKSENEAYQNLLSAEYVPPDDPPKPGDTSEHNPKTPELDGTPI  
GI-50913503 237 NLEATAANKIPSGYRLNIFKSENEAYQNLLSAEYVPPDDPPKPGDTSEHNPKTPELDGTPI

GI-19224134 295 PEDPKHPDESSEPALPPLMPELDGEVPEVPSESLEPALPPLMPELDGEVPEVPSESLE  
GI-50913503 297 PEDPKHPDDNLEPTLPPV-----

GI-19224134 355 PALPPLMPELDGEVPEVPSESLEPALPPLMPELDGEVPEVPSESLEPALPPLMPELDG  
GI-50913503 316 -----LDGEEVPEVPSESLEPALPPLMPELDG

GI-19224134 415 FEVPEKPSVDLPFIEVPRYEFNNKQDQSLAGSGETEVITEVYGNQONPVDIDKKLPNETG  
GI-50913503 343 QEVPEKPSVDLPFIEVPRYEFNNKQDQSLAGSGETEVITEVYGNQONPVDIDKKLPNETG

GI-19224134 475 FSGNMVETEDTKEPEVLMGGQSESVEFTKDTQTGMSGOTTPOVETEDTKEPEVLMGGQSE  
GI-50913503 403 FSGNMVETEDTKEPEVLMGGQSESVEFTKDTQTGMSGOTTPOVETEDTKEPEVLMGGQSE

GI-19224134 535 SVEFTKDTQTGMSGOTTPOVETEDTKEPEVLMGGQSESVEFTKDTQTGMSGOTTPOVETE  
GI-50913503 463 SVEFTKDTQTGMSGOTTPOVETEDTKEPEVLMGGQSESVEFTKDTQTGMSGOTTPOVETE

GI-19224134 595 DTKEPEVLMGGQSESVEFTKDTQTGMSGFSETATIVEDTRPKLVFHFDDNNEPKVEENREK  
GI-50913503 523 DTKEPEVLMGGQSESVEFTKDTQTGMSGFSETATIVEDTRPKLVFHFDDNNEPKVEENREK

GI-19224134 655 PTKNITPILPATGDIENVLAFLGILILSVLSIFSLLNNKQNNKV-  
GI-50913503 583 PTKNITPILPATGDIENVLAFLGILILSVLSIFSLLNNKQNNKV

FIGURE 53

170/487

GI-19745307 1 MTQKNSY FLLSLTGFILGLLLVTIGLSGVSVGHAETRNGANKQGI  
ORF84 WO 2006/078318 NSYN SFLSLTGFILGLLLVTIGLSGVSVGHAETRNGANKQCAI PCT/US2005/027239  
GI-28810263 1 MTQKNSYKLSFLLSLTGFILGLLLVTIGLSGVSVGHAETRNGANKQGAFEIKNN  
GI-21909640 1  
GI-19224141 1 MTQKNSYKLSFLLSLTGFILGLLLVTIGLSGVSVGHAETRNGANKQGSFEIKNNVDQNNKP

GI-19745307 55 ----- KSQEEYNYE  
ORF84 55 ----- KSQEEYNYE  
GI-28810263 55 ----- KSQEEYNYE  
GI-21909640 1 -----  
GI-19224141 61 LPGATFSLTSKDGKGTSVQTFSTNDKGI VDAQNLQPGTYTLKEETAPDGYD KSRITVT

GI-19745307 64 VYDN ----- RNI  
ORF84 64 VYDN ----- RNI  
GI-28810263 64 VYDN ----- RNI  
GI-21909640 1 -----  
GI-19224141 121 VYENG YTKLVENPYNGEII SKAGSKDVSSSLQLENPKMSVVS KYGKTEVSSGAADFY RNI

GI-19745307 71 LDGGEHKLEIKRVDGTGKTYQG FCFQLTNFF TAQGVSKKLYKKLSS  
ORF84 71 LDGGEHKLEIKRVDGTGKTYQG FCFQLTNFF TAQGVSKKLYKKLSS  
GI-28810263 71 LDGGEHKLEIKRVDGTGKTYQG FCFQLTNFF TAQGVSKKLYKKLSS  
GI-21909640 1 MSS  
GI-19224141 181 AAYFKMSF EKKOK E KSETIN E GDT E V L Q L D R L N P K G I S Q D E P I T Y D S A N S P L A I G K Y H

GI-19745307 118 ----- SDEETLK  
ORF84 118 ----- SDEETLK  
GI-28810263 118 ----- SDEETLK  
GI-21909640 4 ----- SDEETLK  
GI-19224141 241 AENHQLIYTFTDYIAGLDKVLQSLAELSLFLENK E V L E N T S I S N F K S T I G G Q E I T Y K G T V N

GI-19745307 125 QYASKYTSNRRGDTSC  
ORF84 125 QYASKYTSNRRGDTSC  
GI-28810263 125 QYASKYTSNRRGDTSC  
GI-21909640 11 QYASKYTSNRRGDTSC  
GI-19224141 301 VLYGNESTKESNYHTNGLSNVCGSIESYNTETGEFVWVYVNP NRTNI PYATMNLWGFR

GI-19745307 141 ----- NLKKQIAKVLTEGYPT  
ORF84 141 ----- NLKKQIAKVLTEGYPT  
GI-28810263 141 ----- NLKKQIAKVLTEGYPT  
GI-21909640 27 ----- NLKKQIAKVLTEGYPT  
GI-19224141 361 ARSNTSDLENDANTSSAELGEIOVYEVPEGEKLPSSYGVDVTKL L R T L I T A G L N G F O M

GI-19745307 157 NKS DWLN GLTENENIEVTQDAIWF  
ORF84 157 NKS DWLN GLTENENIEVTQDAIWF  
GI-28810263 157 NKS DWLN GLTENENIEVTQDAIWF  
GI-21909640 43 NKS DWLN GLTENENIEVTQDAIWF  
GI-19224141 421 TTRORI DFCNNIQNKAFI I KVTGKTQSGKPLVVQSNLASFRGASEYAAFTPVGGN VYEQ

GI-19745307 182 TETTVPADR SYTNRNVNVSQRMKEVYQKLIDTTDID KYEDVQFDLFVPQDTN  
ORF84 182 TETTVPADR SYTNRNVNVSQRMKEVYQKLIDTTDID KYEDVQFDLFVPQDTN  
GI-28810263 182 TETTVPADR SYTNRNVNVSQRMKEVYQKLIDTTDID KYEDVQFDLFVPQDTN  
GI-21909640 68 TETTVPADR SYTNRNVNVSQRMKEVYQKLIDTTDID KYEDVQFDLFVPQDTN  
GI-19224141 481 NEIALSPSKGSGSGKSPETKPSITVANLKRVAQLRFKKMSTDNVPLPEAAFE LRSSNGNS

GI-19745307 233 LQAVISVEPVIESLPITS LKPIAQNDITAKK  
ORF84 233 LQAVISVEPVIESLPITS LKPIAQNDITAKK  
GI-28810263 233 LQAVISVEPVIESLPITS LKPIAQNDITAKK  
GI-21909640 119 LQAVISVEPVIESLPITS LKPIAQNDITAKK  
GI-19224141 541 QKLEASSNTQGEVHFKDLTS GTYDLYETKAPKGYQQVTEKLEATVTVDTTNPAEEMVTWGS

FIGURE 54

171/487

GI-19745307 264 --- **ITVLEKPKPIYFKLYRQLPGEKEVAVDDAELN** **QINS**  
 ORF84 **WO 2006/078318** **ITVLEKPKPIYFKLYRQLPGEKEVAVDDAELN** **QINSEGPCT/US2005/027239**  
 GI-28810263 264 **ITVDAPKEKPIYFKLYRQLPGEKEVAVDDAELN** **QINSEGO**  
 GI-21909640 150 **ITVDAPKEKPIYFKLYRQLPGEKEVAVDDAELN** **QINSEGO**  
 GI-19224141 601 **PHSSVVEANKEVTIANHNETLTFSSCKNIWENDRPDORPAKIQVOLLONGQKMPNQIQEV**

GI-19745307 305 --- **QEISVTWITNQLVTDERGMAYIYSVKEVDK** **NGELLEPKD** **YIKRED**  
 ORF84 305 **QEISVTWITNQLVTDERGMAYIYSVKEVDK** **NGELLEPKD** **YIKRED**  
 GI-28810263 305 **QEISVTWITNQLVTDERGMAYIYSVKEVDK** **NGELLEPKD** **YIKRED**  
 GI-21909640 191 **QEISVTWITNQLVTDERGMAYIYSVKEVDK** **NGELLEPKD** **YIKRED**  
 GI-19224141 661 **TKDNISVYHEKELPKYDARNOEYKYSVEEVN** **VPDGYKVSYLGNDFNTRETEFV** **EQNNF**

GI-19745307 349 **GLTVNTYVVKPTSG** **HYDIEVTFG** **NGHIDITEDTTPDIVSGENQMK**  
 ORF84 349 **GLTVNTYVVKPTSG** **HYDIEVTFG** **NGHIDITEDTTPDIVSGENQMK**  
 GI-28810263 349 **GLTVNTYVVKPTSG** **HYDIEVTFG** **NGHIDITEDTTPDIVSGENQMK**  
 GI-21909640 235 **GLTVNTYVVKPTSG** **HYDIEVTFG** **NGHIDITEDTTPDIVSGENQMK**  
 GI-19224141 721 **NLEFGNAELNGQSGSKIIDEDETLTSTFKGKKIWKNDTABNRPAIQVQLYADGVAVEGQTK**

GI-19745307 394 **QIEGEDS** **KPIDEVT** **ENNLIEFGKNTMPGEE**  
 ORF84 394 **QIEGEDS** **KPIDEVT** **ENNLIEFGKNTMPGEE**  
 GI-28810263 394 **QIEGEDS** **KPIDEVT** **ENNLIEFGKNTMPGEE**  
 GI-21909640 280 **QIEGEDS** **KPIDEVT** **ENNLIEFGKNTMPGEE**  
 GI-19224141 781 **RLSGSGNEWSFEFKNLKKYNGTGNDIISVKEVTVP** **TGYDVTYSANDIINTKREVT** **TOAG**

GI-19745307 424 **DGTNSNKYEEVEDSRPVDTL** **SGLSSEQGS** **SGDMTIEEDSATHIKFSKRD**  
 ORF84 424 **DGTNSNKYEEVEDSRPVDTL** **SGLSSEQGS** **SGDMTIEEDSATHIKFSKRD**  
 GI-28810263 424 **DGTNSNKYEEVEDSRPVDTL** **SGLSSEQGS** **SGDMTIEEDSATHIKFSKRD**  
 GI-21909640 310 **DGTNSNKYEEVEDSRPVDTL** **SGLSSEQGS** **SGDMTIEEDSATHIKFSKRD**  
 GI-19224141 841 **PKLEIETLPL** **ESGASGGTTT** **VEDSRPVDTL** **SGLSSEQGS** **SGDMTIEEDSATHIKFSKRD**

GI-19745307 473 **IDGKELAGATMELRDS** **SGNTISTWISDQVND** **FYLMFGKTYFVETRAPDGYEIATAITFT**  
 ORF84 473 **IDGKELAGATMELRDS** **SGNTISTWISDQVND** **FYLMFGKTYFVETRAPDGYEIATAITFT**  
 GI-28810263 473 **IDGKELAGATMELRDS** **SGNTISTWISDQVND** **FYLMFGKTYFVETRAPDGYEIATAITFT**  
 GI-21909640 359 **IDGKELAGATMELRDS** **SGNTISTWISDQVND** **FYLMFGKTYFVETRAPDGYEIATAITFT**  
 GI-19224141 901 **IDGKELAGATMELRDS** **SGNTISTWISDQVND** **FYLMFGKTYFVETRAPDGYEIATAITFT**

GI-19745307 533 **VNEQGQVTVNGKATIGDAHIVMDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKS**  
 ORF84 533 **VNEQGQVTVNGKATIGDAHIVMDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKS**  
 GI-28810263 533 **VNEQGQVTVNGKATIGDAHIVMDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKS**  
 GI-21909640 419 **VNEQGQVTVNGKATIGDAHIVMDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKS**  
 GI-19224141 961 **VNEQGQVTVNGKATIGDAHIVMDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKS**

GI-19745307 593 **SDVIIGGQG**  
 ORF84 593 **SDVIIGGQG**  
 GI-28810263 593 **SDVIIGGQGEVVDTTEDTQSGMTGHS**  
 GI-21909640 479 **SDVIIGGQGEVVDTTEDTQSGMTGHS** **SGSTTEIEDSKSSDVIIGGQGEVVDTTEDTQSGMT**  
 GI-19224141 1021 **SDLIIGGQGEVVDTTEDTQSGMTGHS**

GI-19745307 602 **QIVETTEDTQTGMHGD** **SGCKTEVEDTKLVQSFHFDNK**  
 ORF84 602 **QIVETTEDTQTGMHGD** **SGCKTEVEDTKLVQSFHFDNK**  
 GI-28810263 619 **GSTTKIEDSKSSDVIIGGQGEVVDTTEDTQTGMHGD** **SGCKTEVEDTKLVQSFHFDNK**  
 GI-21909640 539 **GHS** **GSTTKIEDSKSSDVIIGGQGEVVDTTEDTQTGMHGD** **SGCKTEVEDTKLVQSFHFDNK**  
 GI-19224141 1047 **GSTTKIEDSKSSDVIIGGQGEVVDTTEDTQTGMHGD** **SGCKTEVEDTKLVQSFHFDNK**

GI-19745307 639 **ESESNSSEIPKDKPKSNTSLPATGKQHNMMFFIMVTSCSLISSVFVLSLTKKRLSSC**  
 ORF84 639 **ESESNSSEIPKDKPKSNTSLPATGKQHNMMFFIMVTSCSLISSVFVLSLTKKRLSSC**  
 GI-28810263 676 **EPESNSSEIPKDKPKSNTSLPATGKQHNMMFFIMVTSCSLISSVFVLSLTKKRLSSC**  
 GI-21909640 599 **EPESNSSEIPKDKPKSNTSLPATGKQHNMMFFIMVTSCSLISSVFVLSLTKKRLSSC**  
 GI-19224141 1104 **EPESNSSEIPKDKPKSNTSLPATGKQHNMMFFIMVTSCSLISSVFVLSLTKKRLSSC**

FIGURE 54A



GI-19224WO 2006/078318-MKNNKLLLATAILATALGTASLNQNVKAEAGVSSGQITTKRSITPCT/US2005/027239

ORF80 PCT1-LEBKKMKNNKLLLATAILATALGTASLNQNVKAEAGVVTICKSLQVTKMT-YDDEVLIM  
 GI-21909636 1 MKNNKLLLATAILATALGTASLNQNVKAEAGVSENKLVKNTFDSYTDNEVLM  
 GI-28810259 1 MEREMKNNKLLLATAILATALGTASLNQNVKAEAGVSENKLVKNTFDSYTDNEVLM  
 GI-19745303 1 MKNNKLLLATAILATALGTASLNQNVKAEAGVSDGSILVKKTFPSYTDNEVLM  
 GI-13621428 1 MKLRHLLITG-AALTSFAATT-VHGETVVGAKLTVTNLDLVNSN-ALH

GI-19224137 56 PNTDYTFSVNEDSAATCTESN-LPIKEGIAVN-NOEIKVSVSNTDKTSNGENQVVVDFMK  
 ORF80 60 RETAFTFTLEPDMTASCHEGS-LDIRNGIVBGLDKQVTVKYKNTDKTSOKTILAQFDFSK  
 GI-21909636 56 PRADYTFKVEADSTASGKTGDLGKIKPGIVNGLT-EQILSYTWTDPDSNVKSTEFDFSK  
 GI-28810259 61 PRADYTFKVEADSTASGKTGDLGKIKPGIVNGLT-EQILSYTWTDPDSNVKSTEFDFSK  
 GI-19745303 56 PRADYTFKVEADSTASGKTGDLGKIKPGIVNGLT-EQILSYTWTDPDSNVKSTEFDFSK  
 GI-13621428 48 PNTDYTFKVEADSTASGKTGDLGKIKPGIVNGLT-EQILSYTWTDPDSNVKSTEFDFSK

GI-19224137 114 VTFPSVGYRYRVVTEENKGTAE-GITYDDTKHIVDVYVGN-NEKGGLEPKYIVSKKEDSA  
 ORF80 119 VKFPAICGVYRYMVSEKNDKED-GITYDDTKHIVDVYVGNKANNEEGFEVLIVISKEGESS  
 GI-21909636 115 VVFPGIGVYRYTVSEKNGDVE-GITYDDTKHIVDVYVGN-KEGGGFEPKIVSKKEDSD  
 GI-28810259 120 VVFPGIGVYRYTVSEKNGDVE-GITYDDTKHIVDVYVGN-KEGGGFEPKIVSKKEDSD  
 GI-19745303 116 VKFPGVGVYRYTVSEVNCNFA-GIAYDSQVHTVDVYVGN-REDGGFEAKYIVSTEGGQS  
 GI-13621428 103 VTFEKEGVYRYRYVTEKIDKRVPGVSYDTTSYTVQVHVLWN-EEQQKPVATYIVGYKEGS-

GI-19224137 171 TRPPIQFNNSFETTSLSKTEBVTGNTGDKKRAFTFTLTLOFNEYEASSVVKIEENGQ--  
 ORF80 178 TKKPIQFNNSLSKTTSLKTEKQITGNACDRKSENFLLTLOPSEYYNTGSSVVKIEODGS--  
 GI-21909636 172 VKKPIQFNNSFATTSLSKVKKNVSGNTGELQKEEDFTLTLESTNEKNDQIVSLQKQNE--  
 GI-28810259 177 VKKPIQFNNSFATTSLSKVKKNVSGNTGELQKEEDFTLTLESTNEKNDQIVSLQKQNE--  
 GI-19745303 173 DKKPIQFNNSFEDTTSLSKVKKNVSGNTGELQKEEDFTLTLESTNEKNDQIVSLQKQNE--  
 GI-13621428 161 KVPPIQFNNSLSTTTLVKKKVSCTGCDRSGDENFGLTLKANQYKASEKVMIEKTTKGG

GI-19224137 229 TRDVKIGEAAYKFTLNDSOSVLSKLPVGINYKVEEAEANQGGYTTATLNDG--EK  
 ORF80 236 KNDVIGTPYKFTLGHCKSVMLSKLPVGINYKVEEAEANQGGYTTATLNDG--EK  
 GI-21909636 230 KNDVIGTPYKFTLGHCKSVMLSKLPVGINYKVEEAEANQGGYTTATLNDG--EK  
 GI-28810259 235 KNDVIGTPYKFTLGHCKSVMLSKLPVGINYKVEEAEANQGGYTTATLNDG--EK  
 GI-19745303 231 TRDVKIGEAAYKFTLNDSOSVLSKLPVGINYKVEEAEANQGGYTTATLNDG--EK  
 GI-13621428 220 QAPVTEASIDQLYHFTLNDGESIKVTNLFVGVYVVTEDDYKSEKVTNVEVSPQDGV

GI-19224137 283 LSTYNLG-QEHTLTKTAEIVVTNNRDTQVPTGVVGTLPFAVLISIVAIGGVYITRKK  
 ORF80 292 SSETLSTONOKTDESADAEIVVTNNRDTQVPTGVVGTLPFAVLISIVAIGGVYITRKK  
 GI-21909636 285 SSETLSTONOKTDESADAEIVVTNNRDTQVPTGVVGTLPFAVLISIVAIGGVYITRKK  
 GI-28810259 290 SSETLSTONOKTDESADAEIVVTNNRDTQVPTGVVGTLPFAVLISIVAIGGVYITRKK  
 GI-19745303 285 SSETLSTONOKTDESADAEIVVTNNRDTQVPTGVVGTLPFAVLISIVAIGGVYITRKK  
 GI-13621428 280 KNTAGNSTEQLSTDKDMILFTNNKKEVPTGVVGTLPFAVLISIVAIGGVYITRKK

GI-19224137 342 A  
 ORF80 352 A  
 GI-21909636 344 A  
 GI-28810259 349 A  
 GI-19745303 344 A  
 GI-13621428 340 A

FIGURE 55

173/487

GI-2190WO 2006/078318

PCT/US2005/027239

GI-28810261	1	MLFQSVVMILTMLAFNOTVLAKDSTV
GI-19224139	1	MLFQSVVMILTMLAFNOTVLAKDSTV
ORF82	1	MLFQSVVMILTMLAFNOTVLAKDSTV
GI-19745305	1	MLFQSVVMILTMLAFNOTVLAKDSTV

GI-21909638	20	QTSISVENVLERAGDSTPFPSIALESIDAMKTIEEITAGSGKASFSPLTFTTVGQYTYRV
GI-28810261	26	QTSISVENVLERAGDSTPFPSIALESIDAMKTIEEITAGSGKASFSPLTFTTVGQYTYRV
GI-19224139	26	QTSISVENVLERAGDSTPFPSIALESIDAMKTIEEITAGSGKASFSPLTFTTVGQYTYRV
ORF82	61	QTSISVENVLERAGDSTPFPSIALESIDAMKTIEEITAGSGKASFSPLTFTTVGQYTYRV
GI-19745305	32	QTSISVENVLERAGDSTPFPSIALESIDAMKTIEEITAGSGKASFSPLTFTTVGQYTYRV

GI-21909638	80	YQKPSQNKDYQADTTVFDVLVYVYTYDEDGTLVAKVISRRAGDEEKSAITFKPKRLVKPIP
GI-28810261	86	YQKPSQNKDYQADTTVFDVLVYVYTYDEDGTLVAKVISRRAGDEEKSAITFKPKRLVKPIP
GI-19224139	86	YQKPSQNKDYQADTTVFDVLVYVYTYDEDGTLVAKVISRRAGDEEKSAITFKPKRLVKPIP
ORF82	121	YQKPSQNKDYQADTTVFDVLVYVYTYDEDGTLVAKVISRRAGDEEKSAITFKPKRLVKPIP
GI-19745305	92	YQKPSQNKDYQADTTVFDVLVYVYTYDEDGTLVAKVISRRAGDEEKSAITFKPKRLVKPIP

GI-21909638	140	PRQPNIPKTPPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLSRL
GI-28810261	146	PRQPNIPKTPPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLSRL
GI-19224139	146	PRQPNIPKTPPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLSRL
ORF82	181	PRQPNIPKTPPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLSRL
GI-19745305	152	PRQPNIPKTPPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLSRL

FIGURE 56

PCT/US05/27239 174/487

!!SEQUENCE\_LIST 1.0

(Peptide) FASTA of: gi-50913505.pep from: 1 to: 1036 September 15, 2004 18:46

gi|50913505|ref|YP\_059477.1| Collagen adhesion protein [Streptococcus pyogenes M GAS10394]

TO: \*.pep Sequences: 56 Symbols: 22,803 Word Size: 2

Scoring matrix: GenRunData:blosum50.cmp

Variable pamfactor used

Gap creation penalty: 12 Gap extension penalty: 2

## Histogram Key:

Each histogram symbol represents 1 search set sequences  
z-scores computed from opt scores

z-score	obs	exp
(=)	(*)	
< 20	0	0:
22	0	0:
24	0	0:
26	0	0:
28	0	0:
30	0	0:
32	0	0:
34	0	1:*
36	0	2: *
38	0	3: *
40	0	4: *
42	0	5: *
44	0	5: *
46	5	5:====*
48	10	5:====*=====
50	12	4:====*=====
52	4	4:====*
54	6	3:====*
56	4	3:====*
58	5	2:====*
60	3	2:====*
62	0	1:*
64	1	1:*
66	1	1:*
68	0	1:*
70	1	1:*
72	0	0:
74	0	0:
76	1	0:==
78	0	0:
80	0	0:
82	0	0:
84	0	0:
86	0	0:
88	0	0:
90	0	0:
92	0	0:
94	0	0:
96	1	0:==
98	0	0:

FIGURE 57

PCT/US05/27239

```

100 0 0:
102 0 0:
104 0 0:
106 1 0:=
108 0 0:
110 0 0:
112 0 0:
114 0 0:
116 0 0:
118 0 0:
>120 1 0:=

```

Joining threshold: 39, opt. threshold: 27, opt. width: 16, reg.-scaled

The best scores are:

initl initn opt z-sc E(54)...

```

/home/morاما/gas/pili/align/gi-50913505.pep Begin: 1 End: 1036
! gi|50913505|ref|YP_059477.1| Collag... 6697 6697 6697 3452.1 9.6e-189
/home/morاما/gas/pili/align/gi-19224141.pep Begin: 48 End: 144
! gi|19224141|gb|AAL86412.1|AF447492_... 63 100 159 105.9 0.023
/home/morاما/gas/pili/align/gi-21909640.pep Begin: 147 End: 449
! gi|21909640|ref|NP_663908.1| protei... 35 35 136 96.3 0.08
/home/morاما/gas/pili/align/gi-13621428.pep Begin: 57 End: 318
! gi|13621428|gb|AAK33238.1| hypothet... 33 33 91 75.6 1.1
/home/morاما/gas/pili/align/gi-50913506.pep Begin: 33 End: 428
! gi|50913506|ref|YP_059478.1| Fimbri... 70 149 86 71.3 1.9
/home/morاما/gas/pili/align/gi-13621432.pep Begin: 14 End: 56
! gi|13621432|gb|AAK33241.1| conserve... 40 65 78 68.0 2.9
/home/morاما/gas/pili/align/gi-19745301.pep Begin: 241 End: 466
! gi|19745301|ref|NP_606437.1| putati... 52 52 73 64.8 4.3
/home/morاما/gas/pili/align/gas15.pep Begin: 492 End: 739
! GAS15 GAS15 43 68 69 61.4 6.6
/home/morاما/gas/pili/align/gi-21909636.pep Begin: 176 End: 298
! gi|21909636|ref|NP_663904.1| conser... 31 31 62 60.8 7.1
/home/morاما/gas/pili/align/gi-28810259.pep Begin: 181 End: 303
! gi|28810259|dbj|BAC63197.1| hypothe... 31 31 62 60.7 7.2
/home/morاما/gas/pili/align/gi-19224139.pep Begin: 90 End: 143
! gi|19224139|gb|AAL86410.1|AF447492_... 43 43 54 58.9 8.9
/home/morاما/gas/pili/align/gi-19745305.pep Begin: 96 End: 149
! gi|19745305|ref|NP_606441.1| hypoth... 43 43 54 58.8 9
/home/morاما/gas/pili/align/orf82.pep Begin: 125 End: 178
! TRANSLATE of: orf82.seq check: 4296... 43 43 54 58.2 9.6
/home/morاما/gas/pili/align/gi-21909638.pep Begin: 84 End: 137
! gi|21909638|ref|NP_663906.1| hypoth... 43 43 52 58.0 9.9
\\End of List

```

gi-50913505.pep

/home/morاما/gas/pili/align/gi-50913505.pep

gi|50913505|ref|YP\_059477.1| Collagen adhesion protein [Streptococcus pyogenes M GAS10394]

SCORES Initl: 6697 Initn: 6697 Opt: 6697 z-score: 3452.1 E(): 9.6e-189  
 >>/home/morاما/gas/pili/align/gi-50913505.pep (1036 aa)  
 initn: 6697 initl: 6697 opt: 6697 Z-score: 3452.1 expect(): 9.6e-189  
 Smith-Waterman score: 6697; 100.0% identity in 1036 aa overlap  
 (1-1036:1-1036)

10 20 30 40 50 60  
 gi-50913505. MYSRLKRELIVIVINRKKKYKLIRLMVTVGLIFSQVLPLIRRLGLQMIQTQKVIPQEIQT

FIGURE 57A

PC T11477455.27.23.33

gi-50913505. MYSRLKRELIVIVINRKKKYKILRLMVTVGLIFSQVLVPIRRLGLQMISTQTKVIPQEI VT  
10 20 30 40 50 60

70 80 90 100 110 120  
gi-50913505. QTETQGTQVVATKQKLESENSSLKVALKRESGFEHNATIDASLDTESQGDNSQRSVTQAI  
|||||  
gi-50913505. QTETQGTQVVATKQKLESENSSLKVALKRESGFEHNATIDASLDTESQGDNSQRSVTQAI  
70 80 90 100 110 120

130 140 150 160 170 180  
gi-50913505. VTMALELRKQGLSIVDTKIVRIQSSTNQRNDITTTLTTFKNGLSLEGASTEANDPNVRVGI  
|||||  
gi-50913505. VTMALELRKQGLSIVDTKIVRIQSSTNQRNDITTTLTTFKNGLSLEGASTEANDPNVRVGI  
130 140 150 160 170 180

190 200 210 220 230 240  
gi-50913505. VNPNDTVQTITPTIKQDADGKVKNLVFTGRLGKQVIIIVSTTRLKEEQTISLDSYGELVID  
|||||  
gi-50913505. VNPNDTVQTITPTIKQDADGKVKNLVFTGRLGKQVIIIVSTTRLKEEQTISLDSYGELVID  
190 200 210 220 230 240

250 260 270 280 290 300  
gi-50913505. GAVGLSQKDRPPYSKPI TVNILKPKLSSIESSLDSKDFEIVKTIDNLYTWDDQFYLLDFI  
|||||  
gi-50913505. GAVGLSQKDRPPYSKPI TVNILKPKLSSIESSLDSKDFEIVKTIDNLYTWDDQFYLLDFI  
250 260 270 280 290 300

310 320 330 340 350 360  
gi-50913505. SKQYEVLKTDYQSAKDSTPQTRDILFGEYTV EPLVMNKGHNNTINIYRSTRPLGLKPIG  
|||||  
gi-50913505. SKQYEVLKTDYQSAKDSTPQTRDILFGEYTV EPLVMNKGHNNTINIYRSTRPLGLKPIG  
310 320 330 340 350 360

370 380 390 400 410 420  
gi-50913505. AAPALIQPRSFRLTPRSTRMKRSAPVEKFEGELEHHKRIDYLGDNQNNPD TTIDDKED E  
|||||  
gi-50913505. AAPALIQPRSFRLTPRSTRMKRSAPVEKFEGELEHHKRIDYLGDNQNNPD TTIDDKED E  
370 380 390 400 410 420

430 440 450 460 470 480  
gi-50913505. HDTSDLYRLYLDMTGKKNP LDILVVVDKSGSMQEGIGSVQRYRYAQRWDDYYSQWVYHG  
|||||  
gi-50913505. HDTSDLYRLYLDMTGKKNP LDILVVVDKSGSMQEGIGSVQRYRYAQRWDDYYSQWVYHG  
430 440 450 460 470 480

490 500 510 520 530 540  
gi-50913505. TFDYSSYQGESFNRGQIHYRYRGIVSVSDGIRRDDAVKNSLLGVNGLLQRFVNINPENKL  
|||||  
gi-50913505. TFDYSSYQGESFNRGQIHYRYRGIVSVSDGIRRDDAVKNSLLGVNGLLQRFVNINPENKL  
490 500 510 520 530 540

550 560 570 580 590 600  
gi-50913505. SVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNSRDAELLKGWSTNSLLDPNTLTALHNN  
|||||  
gi-50913505. SVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNSRDAELLKGWSTNSLLDPNTLTALHNN  
550 560 570 580 590 600

610 620 630 640 650 660  
gi-50913505. GTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPTFFYFGEDGYRSGNGSSNDRNNVTRS  
|||||  
gi-50913505. GTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPTFFYFGEDGYRSGNGSSNDRNNVTRS  
610 620 630 640 650 660

FIGURE 57B

PCT/US05/27239

```

      670      680      690      700      710      720
gi-50913505. QEGSKLAIDEFKARYPNLSIYSLGVSKDINSDTASSSPVVLKYLSGEEHYGITDTAELE
|||||
gi-50913505. QEGSKLAIDEFKARYPNLSIYSLGVSKDINSDTASSSPVVLKYLSGEEHYGITDTAELE
      670      680      690      700      710      720

      730      740      750      760      770      780
gi-50913505. KTLNKIVEDSKLSQLGISDSLSQYVDYYDKQPDVLVTRKSKVNDETEILYQKDQVQEAGK
|||||
gi-50913505. KTLNKIVEDSKLSQLGISDSLSQYVDYYDKQPDVLVTRKSKVNDETEILYQKDQVQEAGK
      730      740      750      760      770      780

      790      800      810      820      830      840
gi-50913505. DIIDKVVFTPKTTSQPKGKVTLTFFKSDYKVDDEYTYTSLFNVKASDEAYEKYKDNREGYS
|||||
gi-50913505. DIIDKVVFTPKTTSQPKGKVTLTFFKSDYKVDDEYTYTSLFNVKASDEAYEKYKDNREGYS
      790      800      810      820      830      840

      850      860      870      880      890      900
gi-50913505. EMGDSDDTYGTNQTSSGKGLPSNSDASVNYMADGREQKLPYKHPVIQVKTVPITFTKVD
|||||
gi-50913505. EMGDSDDTYGTNQTSSGKGLPSNSDASVNYMADGREQKLPYKHPVIQVKTVPITFTKVD
      850      860      870      880      890      900

      910      920      930      940      950      960
gi-50913505. ADNNQKKLAGVEFELRKEDKKIVWEKGTTSNGQLNFKYLQKGKTYLYETKAKLGYTLF
|||||
gi-50913505. ADNNQKKLAGVEFELRKEDKKIVWEKGTTSNGQLNFKYLQKGKTYLYETKAKLGYTLF
      910      920      930      940      950      960

      970      980      990      1000      1010      1020
gi-50913505. ENPWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNYKIYQLPSSGGRGSQIFIIVGSMTA
|||||
gi-50913505. ENPWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNYKIYQLPSSGGRGSQIFIIVGSMTA
      970      980      990      1000      1010      1020

      1030
gi-50913505. TVALLFYRRQHRKKQY
|||||
gi-50913505. TVALLFYRRQHRKKQY
      1030

gi-50913505.pep
/home/morana/gas/pili/align/gi-19224141.pep
gi|19224141|gb|AAL86412.1|AF447492.9 protein F2 [Streptococcus pyogenes]

SCORES   Init1: 63   Initn: 100   Opt: 159   z-score: 105.9 E(): 0.023
>>/home/morana/gas/pili/align/gi-19224141.pep (1161 aa)
initn: 100 init1: 63 opt: 159 Z-score: 105.9 expect(): 0.023
Smith-Waterman score: 159; 36.7% identity in 98 aa overlap
(895-990:48-144)

      870      880      890      900      910      920
gi-50913505. SDASVNYMADGREQKLPYKHPVIQVKTVPITFTKVDADNNQKKLAGVEFELRKEDKK-IV
      :| :|:| :| :| :|
gi-19224141. FILGLLLVFIGLSGVSVGHAETRNGANKQGSFEIKKVDQNNKPLPGATFSLTSKDGKGT
      20      30      40      50      60      70

      930      940      950      960      970      980

```

FIGURE 57C

FIGURE 57D



FIGURE 57E

**FIGURE 57F**

PCT/US05/27239 181/487

s1

SCORES Init1: 40 Initn: 65 Opt: 78 z-score: 68.0 E(): 2.9  
 >>/home/morama/gas/pili/align/gi-13621432.pep (450 aa)  
 initn: 65 init1: 40 opt: 78 Z-score: 68.0 expect(): 2.9  
 Smith-Waterman score: 78; 37.0% identity in 46 aa overlap  
 (368-411:14-56)

```

      340      350      360      370      380      390
gi-50913505. KGHNNTINIYIRSTRPLGLKPIGAAPALIQPRSFRLTPRSTRMK--RSAPVEKFECELE
                  |  ::| |  ::| |  ::| |  ::| |
gi-13621432.      MTRTNYQKKRMTCPVETEDITYRRKKIKGRRQAILAQFEPELV
                  10      20      30      40

      400      410      420      430      440      450
gi-50913505. HHKRIDYLGDNQNNPDTTIDDKEDHDTSDLYRLYLDMTGKKNPLDIEVVVDKSGSMQEG
      ||: |  ||: : ||
gi-13621432. HHELI---GDSCTCPDCHGTLTEIGSVVQROELVFIPAQLKRINHVOHAYKCQTCSDNSL
                  50      60      70      80      90      100

```

gi-50913505.pep

/home/morama/gas/pili/align/gi-19745301.pep

gi|19745301|ref|NP\_606437.1| putative collagen binding protein [Streptococcus pyogenes MGAS8232]

SCORES Init1: 52 Initn: 52 Opt: 73 z-score: 64.8 E(): 4.3  
 >>/home/morama/gas/pili/align/gi-19745301.pep (524 aa)  
 initn: 52 init1: 52 opt: 73 Z-score: 64.8 expect(): 4.3  
 Smith-Waterman score: 95; 23.7% identity in 245 aa overlap  
 (759-989:241-466)

```

      730      740      750      760      770      780
gi-50913505. DSKLSQLGISDSLSQYVDYYDKQPDVLVTRKSKVNDETEILYQKDQVQVQAG-KDIIDKVV
                  ::::||| :| | :| | :| |:::
gi-19745301. ETIDPDFNEGKEIKYTHILGADLFSYANNPRASTNDE--LLSQVKVLEKGYRD--DSTT
                  220      230      240      250      260

      790      800      810      820      830      840
gi-50913505. FTPKTTSPQPKGVTLT---FKSDYKVD--EYTYLSEFNVKASDEAYEKYKDNENGRYSEM
      :: |:: :|:: :|:: :|:: :|:: :|:: :|::
gi-19745301. YANLTSVEFRAATGLAIYYFTDSVDLDNLADYHGFGLTTEALNATKEIVAYAEADRANLP
      270      280      290      300      310      320

      850      860      870      880      890
gi-50913505. GDSDTDY---GTNQTSS--GKGGLPSNSDASVNYMADGREQKLPHYKHPVIQVKTIVPITFT
      : | :| :| :| :| :| :| :| :| :| :| :| :|
gi-19745301. NISNLDIFYVPNSNKYQSLIGTQYHP-ESLVDIIRMEDKQAPIIPITHKLTISKTVTGTI-
      330      340      350      360      370      380

      900      910      920      930      940      950
gi-50913505. KVDADNNQKKLAGVEFELRKEDKKIVWEKGTGSGN-GQLNFKYLQKQK-TYYLYETKAKL
      || :|| :|:::| : : :|| :| :| :| :| :| :|
gi-19745301. ---AD--KKKEFNFEIHLKSSDQAI--SGTYPTNSGELT---VTDGKATFTLKDGESLI
                  390      400      410      420      430

      960      970      980      990      1000      1010
gi-50913505. GYTLPEN-PWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNYKIYQLPSSGGRGSQIFII
      || :| :|:: :| :| :| :| :| :| :| :| :|
gi-19745301. VEGLPSPGYSYEITETGASDYEVS--VNGK-NAPDGKATKASVKEDETITFENRKDLVPPT

```

FIGURE 57G

PCT/US05/27239

440

450

460

470

480

490

1020

1030

gi-50913505. VGSMTATVALLFYRRQHRKKQY

gi-19745301. GLTTDGAIIYLWLLLLLVLLGLWVWLIGRKLKND

500

510

520

gi-50913505.pep

/home/morama/gas/pili/align/gas15.pep

GAS15 GAS15

SCORES Init1: 43 Initn: 68 Opt: 69 z-score: 61.4 E(): 6.6

&gt;&gt;/home/morama/gas/pili/align/gas15.pep (762 aa)

initn: 68 init1: 43 opt: 69 z-score: 61.4 expect(): 6.6

Smith-Waterman score: 100; 21.4% identity in 252 aa overlap

(641-873:492-739)

```

          620      630      640      650      660
gi-50913505. AKEILNEVKDDGRRKIMIFISDGVPTFYFGEDGYRSGNGSSNDRNNVTRSQ--EGSKLAI
          |||| :|:: : :|::| :|::| :|::|
gas15.pep    HIAGRDLFKYTVKPRDTPDPTFLKHIKKVIEKGYRE-KGQAI EYSGLTETQLRAATQLAI
          470      480      490      500      510      520

```

```

        670      680      690      700      710      720
gi-50913505. DEF--KARYPNLSIYSLGVSKDINSDTASSPVVLKYLSGEEHYGYITDTAELEKTLNKI
          |::: : : : :|:::| : : : :|:::| : : : :|:::|
gas15.pep    YYFTDSAE LDKDKLDYHGF GDMNDSTLAVAKILVEY-AQDSNPPQLTDLDFIPNNNKY
          530      540      550      560      570

```

```

        730      740      750      760      770
gi-50913505. VEDSKLSQLGISDSLSQYVDYYDKQPDVLVT----RKSQVN---DETEILYQKDQVQEA
          :| : : : :| : : : :| : : : :| : : : :| : : : :|
gas15.pep    --QSLIGTQWHPEDLVDIIRMEDKKEVIPVTHNLTLRKTVTGLAGDRTKDFHFEIELKNN
          580      590      600      610      620      630

```

```

        780      790      800      810      820      830
gi-50913505. GKDIIDKVVF TPKTTSQPK-GKVTLTFSKSDYKVDDE-YTYTLSFNVKASDEAYEKYKDNE
          : : : : :| :| : :| :| : :| : : : :| : : : :| :| :|
gas15.pep    KOELLSQTVKTDKTNLEFKDGKATINLKHGESLTLOGLPEGYSYLVKETDSEGYKVKVNS
          640      650      660      670      680      690

```

```

        840      850      860      870      880      890
gi-50913505. GRYSMGDSDDTDYGTNOT----SSGKGGLPSNSDASVN-YMADGREOKLPYKHPVIOVKI
          : : : :| :| : :| : : : :| : : : :| : : : :| :| :|
gas15.pep    QEVANATVSKTGITSDETLAFENNKPEVVPVTVGVQDKINGYLALIVIAGISLGIWGIHTIR
          700      710      720      730      740      750

```

```

        900      910      920      930      940      950
gi-50913505. VPITFTKVDADNNQKKLAGVEFELRKEDKKIVWEKGTGSGNQLNFKYLQKGKTYLYET
          |||||
gas15.pep    IRKHD
          760

```

gi-50913505.pep

/home/morama/gas/pili/align/gi-21909636.pep

gi|21909636|ref|NP\_663904.1| conserved hypothetical protein [Streptococcus pyogenes MGAS315]

FIGURE 57H

183/487

SCORES Init1: 31 Initn: 31 Opt: 62 z-score: 60.8 E(): 7.1  
 >>/home/morama/gas/pili/align/gi-21909636.pep (344 aa)  
 initn: 31 init1: 31 opt: 62 Z-score: 60.8 expect(): 7.1  
 Smith-Waterman score: 71; 22.9% identity in 131 aa overlap  
 (181-305:176-298)

```

      160      170      180      190      200      210
gi-50913505. DITTTTLTFKNGLSLEGASTEANDPNVRVGVINPNDTVQTTPTIKQDADGKVKNLVFTGR
      || |:: | : :|:::|: ||:
gi-21909636. VDVYVGNKEGGGFEPKFIVSKEQGTDVKKPVNFNNSFATTSLKVKKNVSGN-----TGE
      150      160      170      180      190

      220      230      240      250      260
gi-50913505. LGKQ----VIIVSTTRLKEEQTISLDSYGELVIDGAVGLSQKDRPPYSKPITVNLKPKL
      | | : : : :| :|::| :||::| : : :| : : :| : : :| : :
gi-21909636. LQKEFDFTLTINESTNFKKQIVSLQK-GNEKFEVKIGTPYKFKLNKNGESIQLDKLPVGI
      200      210      220      230      240      250

      270      280      290      300      310      320
gi-50913505. SSISSLDSKDFEIVKTIIDNLYTWDDQ--FYLLDFISKQYEVLTQYQSAKDSTPQTRDI
      : : ::: : || :| || :| ||: :| |
gi-21909636. TYKVNEMEANK-DGYKTTASLKEGDGQSKMYQLDMEQKTDESADIVVTNKRDTQVPTGV
      260      270      280      290      300      310

      330      340      350      360      370      380
gi-50913505. LFGEYTVPEPLVMNKGHNNTINIYIRSTRLGLKPIGAAPALIQPRSFRLTPRSTRMKRS
gi-21909636. VGTLPAPFAVLSIVAIGGVIIYITKRKKA
      320      330      340

```

gi-50913505.pep  
 /home/morama/gas/pili/align/gi-28810259.pep

gi|28810259|dbj|BAC63197.1| hypothetical protein [Streptococcus pyogenes SSI-1]

SCORES Init1: 31 Initn: 31 Opt: 62 z-score: 60.7 E(): 7.2  
 >>/home/morama/gas/pili/align/gi-28810259.pep (349 aa)  
 initn: 31 init1: 31 opt: 62 Z-score: 60.7 expect(): 7.2  
 Smith-Waterman score: 71; 22.9% identity in 131 aa overlap  
 (181-305:181-303)

```

      160      170      180      190      200      210
gi-50913505. DITTTTLTFKNGLSLEGASTEANDPNVRVGVINPNDTVQTTPTIKQDADGKVKNLVFTGR
      || |:: | : :|:::|: ||:
gi-28810259. VDVYVGNKEGGGFEPKFIVSKEQGTDVKKPVNFNNSFATTSLKVKKNVSGN-----TGE
      160      170      180      190      200

      220      230      240      250      260
gi-50913505. LGKQ----VIIVSTTRLKEEQTISLDSYGELVIDGAVGLSQKDRPPYSKPITVNLKPKL
      | | : : : :| :|::| :||::| : : :| : : :| : : :| : :
gi-28810259. LQKEFDFTLTINESTNFKKQIVSLQK-GNEKFEVKIGTPYKFKLNKNGESIQLDKLPVGI
      210      220      230      240      250      260

      270      280      290      300      310      320
gi-50913505. SSISSLDSKDFEIVKTIIDNLYTWDDQ--FYLLDFISKQYEVLTQYQSAKDSTPQTRDI
      : : ::: : || :| || :| ||: :| |
gi-28810259. TYKVNEMEANK-DGYKTTASLKEGDGQSKMYQLDMEQKTDESADIVVTNKRDTQVPTGV
      270      280      290      300      310      320

```

FIGURE 57I



PCT/US05/27239

330 340 350 360 370 380  
 gi-50913505. LFGEYTVPEPLVMNKGHNNTINIIYIRSTRPIGLKPIGAAPALIQPRSFRLTPRSTRMKRS

gi-28810259. VGTLPAPFAVLSIVAIGGVIIYITKRKKA  
 330 340

gi-50913505.pep  
 /home/morاما/gas/pili/align/gi-19224139.pep

gi|19224139|gb|AAL86410.1|AF447492\_7 unknown [Streptococcus pyogenes]

SCORES Init1: 43 Initn: 43 Opt: 54 z-score: 58.9 E(): 8.9  
 >>/home/morاما/gas/pili/align/gi-19224139.pep (189 aa)  
 initn: 43 init1: 43 opt: 54 Z-score: 58.9 expect(): 8.9  
 Smith-Waterman score: 54; 31.6% identity in 57 aa overlap  
 (742-796:90-143)

	720	730	740	750	760
gi-50913505.	GITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYDKQP--DVLVTRKSKVNDTEIL				
					:
gi-19224139.	ITIAGSGKASFSPITFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYV---TYDEDGTL				
	60	70	80	90	100 110

	770	780	790	800	810	820
gi-50913505.	YQKDQVQEAGKDIIDKVVFPTKTTSPQKGVTLTFKSDYKVDDEYTYTSLFNVKASDEAY					
		::   :	: ::     :			
gi-19224139.	VAKVISRRAGDEEKSAITFKPKRLVKPIPPROPNIPKTPPLAGEVKSLLGILSIVLLGL					
	120	130	140	150	160	170

gi-50913505.pep  
 /home/morاما/gas/pili/align/gi-19745305.pep

gi|19745305|ref|NP\_606441.1| hypothetical protein [Streptococcus pyogenes MGAS82 32]

SCORES Init1: 43 Initn: 43 Opt: 54 z-score: 58.8 E(): 9  
 >>/home/morاما/gas/pili/align/gi-19745305.pep (195 aa)  
 initn: 43 init1: 43 opt: 54 Z-score: 58.8 expect(): 9  
 Smith-Waterman score: 54; 31.6% identity in 57 aa overlap  
 (742-796:96-149)

	720	730	740	750	760
gi-50913505.	GITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYDKQP--DVLVTRKSKVNDTEIL				
					:
gi-19745305.	ITIAGSGKASFSPITFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYV---TYDEDGTL				
	70	80	90	100	110 120

	770	780	790	800	810	820
gi-50913505.	YQKDQVQEAGKDIIDKVVFPTKTTSPQKGVTLTFKSDYKVDDEYTYTSLFNVKASDEAY					
		::   :	: ::     :			
gi-19745305.	VAKVISRRAGDEEKSAITFKPKRLVKPIPPROPDIPKTPPLAGEVKSLLGILSIVLLGL					
	130	140	150	160	170	180

gi-50913505.pep  
 /home/morاما/gas/pili/align/orf82.pep

TRANSLATE of: orf82.seq check: 4296 from: 1 to: 672  
 generated symbols 1 to: 224.

FIGURE 57J

PCT/US2005/027239  
GETSEQ from morama, September 13, 2004 17:09.

SCORES Init1: 43 Initn: 43 Opt: 54 z-score: 58.2 E(): 9.6  
>>/home/morama/gas/pili/align/orf82.pep (224 aa)  
initn: 43 init1: 43 opt: 54 Z-score: 58.2 expect(): 9.6  
Smith-Waterman score: 54; 31.6% identity in 57 aa overlap  
(742-796:125-178)

	720	730	740	750	760
gi-50913505.	GITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYYDKQP--DVLVTRKSKVNDETEIL				
orf82.pep	ITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYV---TYDEDGTL				
	100	110	120	130	140 150

	770	780	790	800	810	820
gi-50913505.	YQKDQVQEAGKDIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTLSFNVKASDEAY					
		::   :	: ::	:		
orf82.pep	VAKVISRRAGDEEKSAITFKPKRLVKPIPPRQPNIPKTPPLAGEVKSLLGILSIVLLGL					
	160	170	180	190	200	210

gi-50913505.pep  
/home/morama/gas/pili/align/gi-21909638.pep

gi|21909638|ref|NP\_663906.1| hypothetical protein [Streptococcus pyogenes MGAS315]

SCORES Init1: 43 Initn: 43 Opt: 52 z-score: 58.0 E(): 9.9  
>>/home/morama/gas/pili/align/gi-21909638.pep (183 aa)  
initn: 43 init1: 43 opt: 52 Z-score: 58.0 expect(): 9.9  
Smith-Waterman score: 52; 31.6% identity in 57 aa overlap  
(742-796:84-137)

	720	730	740	750	760
gi-50913505.	GITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYYDKQP--DVLVTRKSKVNDETEIL				
gi-21909638.	ITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYV---TYDEDGTL				
	60	70	80	90	100 110

	770	780	790	800	810	820
gi-50913505.	YQKDQVQEAGKDIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTLSFNVKASDEAY					
		::   :	: ::	:		
gi-21909638.	VAKVISRRAGDEEKSAITFKPKRLVKPIPPRQPNIPKTPPLAGEVKSLLGILSIVLLGL					
	120	130	140	150	160	170

! Distributed over 1 thread.

! Start time: Wed Sep 15 18:46:18 2004

! Completion time: Wed Sep 15 18:46:21 2004

! CPU time used:

! Database scan: 0:00:00.1

! Post-scan processing: 0:00:02.2

! Total CPU time: 0:00:02.3

! Output File: gi-50913505.fasta

FIGURE 57K

PCT/US2005/027239  
 !!SEQUENCE\_LIST 1.0

(Peptide) FASTA of: gi-50913506.pep from: 1 to: 556 September 15, 2004 18:45

gi|50913506|ref|YP\_059478.1| Fimbrial structural subunit [Streptococcus pyogenes  
 MGAS10394]

TO: \*.pep Sequences: 56 Symbols: 22,803 Word Size: 2

Scoring matrix: GenRunData:blosum50.cmp

Variable pamfactor used

Gap creation penalty: 12 Gap extension penalty: 2

#### Histogram Key:

Each histogram symbol represents 1 search set sequences  
 z-scores computed from opt scores

z-score	obs	exp
(=)	(*)	
< 20	0	0:
22	0	0:
24	0	0:
26	0	0:
28	0	0:
30	0	0:
32	0	0:
34	0	1:*
36	0	2: *
38	0	3: *
40	0	4: *
42	0	5: *
44	0	5: *
46	1	5:= *
48	8	5:=====*
50	24	5:=====*
52	2	4:== *
54	3	3:==*
56	1	3:= *
58	2	2:==*
60	3	2:==*
62	0	2: *
64	2	1:*=
66	2	1:*=
68	1	1:*
70	2	1:*=
72	0	0:
74	3	0:==
76	1	0:=
78	0	0:
80	0	0:
82	0	0:
84	0	0:
86	0	0:
88	0	0:
90	0	0:
92	0	0:
94	0	0:
96	0	0:
98	0	0:

FIGURE 58

```

100      0      0:
102      0      0:
104      0      0:
106      0      0:
108      0      0:
110      0      0:
112      0      0:
114      0      0:
116      0      0:
118      0      0:
>120     1      0:=

```

Joining threshold: 37, opt. threshold: 25, opt. width: 16, reg.-scaled

The best scores are:

initl initn opt z-sc E(55)...

```

/home/morana/gas/pili/align/gi-50913506.pep      Begin: 1 End: 556
! gi|50913506|ref|YP_059478.1| Fimbri... 3454 3454 3454 1016.2 4.7e-53
/home/morana/gas/pili/align/orf84.pep      Begin: 316 End: 567
! TRANSLATE of: orf84.seq check: 7868... 57 83 135 75.1 1.2
/home/morana/gas/pili/align/gi-19745307.pep      Begin: 316 End: 567
! gi|19745307|ref|NP_606443.1| protei... 57 83 135 75.1 1.2
/home/morana/gas/pili/align/gi-21909640.pep      Begin: 202 End: 524
! gi|21909640|ref|NP_663908.1| protei... 56 81 134 75.0 1.2
/home/morana/gas/pili/align/gi-28810263.pep      Begin: 316 End: 638
! gi|28810263|dbj|BAC63201.1| protein... 56 82 134 74.7 1.3
/home/morana/gas/pili/align/orf80.pep      Begin: 49 End: 352
! TRANSLATE of: orf80.seq check: 9824... 45 69 113 70.8 2.1
/home/morana/gas/pili/align/gi-19224137.pep      Begin: 25 End: 342
! gi|19224137|gb|AAL86408.1|AF447492... 45 69 109 69.8 2.4
/home/morana/gas/pili/align/gi-19224141.pep      Begin: 277 End: 645
! gi|19224141|gb|AAL86412.1|AF447492... 73 73 118 68.9 2.7
/home/morana/gas/pili/align/gi-21909636.pep      Begin: 44 End: 344
! gi|21909636|ref|NP_663904.1| conser... 45 98 96 66.1 3.8
/home/morana/gas/pili/align/gi-28810259.pep      Begin: 49 End: 349
! gi|28810259|dbj|BAC63197.1| hypothe... 45 98 96 66.0 3.8
/home/morana/gas/pili/align/gas15.pep      Begin: 222 End: 470
! GAS15 GAS15... 42 68 96 63.8 5
/home/morana/gas/pili/align/gi-13621428.pep      Begin: 17 End: 340
! gi|13621428|gb|AAK33238.1| hypothet... 41 41 87 63.6 5.2
/home/morana/gas/pili/align/gi-19224135.pep      Begin: 193 End: 462
! gi|19224135|gb|AAL86406.1|AF447492... 41 41 86 61.0 7
/home/morana/gas/pili/align/gi-50913505.pep      Begin: 503 End: 966
! gi|50913505|ref|YP_059477.1| Collag... 70 149 86 60.1 7.8
/home/morana/gas/pili/align/gi-13621430.pep      Begin: 60 End: 143
! gi|13621430|gb|AAK33240.1| hypothet... 43 67 67 59.2 8.7
/home/morana/gas/pili/align/gi-19745303.pep      Begin: 44 End: 344
! gi|19745303|ref|NP_606439.1| hypoth... 51 106 69 58.4 9.5
\\End of List

```

gi-50913506.pep

/home/morana/gas/pili/align/gi-50913506.pep

gi|50913506|ref|YP\_059478.1| Fimbrial structural subunit [Streptococcus pyogenes MGAS10394]

SCORES Initl: 3454 Inith: 3454 Opt: 3454 z-score: 1016.2 E(): 4.7e-53

>>/home/morana/gas/pili/align/gi-50913506.pep (556 aa)

initn: 3454 initl: 3454 opt: 3454 Z-score: 1016.2 expect(): 4.7e-53

Smith-Waterman score: 3454; 100.0% identity in 556 aa overlap

FIGURE 58A

PCT/US2005/027239

(1-556:1-556)

	10	20	30	40	50	60
gi-50913506.	MTNRRET	VREKILITAKK	LMACLA	LAILAVVGLGM	TRVSALS	SKDDTAQLKITNIEGGPTVT
gi-50913506.	MTNRRET	VREKILITAKK	LMACLA	LAILAVVGLGM	TRVSALS	SKDDTAQLKITNIEGGPTVT
	10	20	30	40	50	60
	70	80	90	100	110	120
gi-50913506.	LYKIGEGV	YNTNGDSFIN	KYAEGVSLT	TETGPTSQ	EITTIANG	INTGKIKPFSTENV
gi-50913506.	LYKIGEGV	YNTNGDSFIN	KYAEGVSLT	TETGPTSQ	EITTIANG	INTGKIKPFSTENV
	70	80	90	100	110	120
	130	140	150	160	170	180
gi-50913506.	NGTATYN	NARGASVYI	ALLTGATD	GRTYNPIL	LAASYN	GEENLVTKNIDSKSNYLYGQTSV
gi-50913506.	NGTATYN	NARGASVYI	ALLTGATD	GRTYNPIL	LAASYN	GEENLVTKNIDSKSNYLYGQTSV
	130	140	150	160	170	180
	190	200	210	220	230	240
gi-50913506.	AKSSSLP	SITKKVGT	IDDVNKK	TTSLGSLV	SYSLTFEL	PSYTK
gi-50913506.	AKSSSLP	SITKKVGT	IDDVNKK	TTSLGSLV	SYSLTFEL	PSYTK
	190	200	210	220	230	240
	250	260	270	280	290	300
gi-50913506.	TFNFNSL	TVEWKGM	ANIT	EDGSVM	VENTKIGI	AKEVNNGFNLSFIYDSLESISPNI
gi-50913506.	TFNFNSL	TVEWKGM	ANIT	EDGSVM	VENTKIGI	AKEVNNGFNLSFIYDSLESISPNI
	250	260	270	280	290	300
	310	320	330	340	350	360
gi-50913506.	AVVNNKA	IVGEEGN	PNKAEFF	YSNNPTK	GN	TYDNLDKKPKDGN
gi-50913506.	AVVNNKA	IVGEEGN	PNKAEFF	YSNNPTK	GN	TYDNLDKKPKDGN
	310	320	330	340	350	360
	370	380	390	400	410	420
gi-50913506.	FRKVD	SVSKT	PLIGAIF	GVYDTSN	KLIDIV	TNKN
gi-50913506.	FRKVD	SVSKT	PLIGAIF	GVYDTSN	KLIDIV	TNKN
	370	380	390	400	410	420
	430	440	450	460	470	480
gi-50913506.	LNTE	TYEITAN	WVTATV	KTSANSK	STTYTSD	KNKATDN
gi-50913506.	LNTE	TYEITAN	WVTATV	KTSANSK	STTYTSD	KNKATDN
	430	440	450	460	470	480
	490	500	510	520	530	540
gi-50913506.	DVKEAY	IESTKAL	TGTTFS	KSNEGS	GTVLLET	DIPNTKL
gi-50913506.	DVKEAY	IESTKAL	TGTTFS	KSNEGS	GTVLLET	DIPNTKL
	490	500	510	520	530	540
	550					
gi-50913506.	AMIGA	IGIYIV	KRRKA			
gi-50913506.	AMIGA	IGIYIV	KRRKA			
	550					

FIGURE 58B



189/487  
PCT/US05/27239  
gi-50913506.pep  
/home/morama/gas/pili/align/orf84.pep

TRANSLATE of: orf84.seq check: 7868 from: 1 to: 2088  
generated symbols 1 to: 696.  
GETSEQ from morama, September 13, 2004 17:07.

SCORES Init1: 57 Initn: 83 Opt: 135 z-score: 75.1 E(): 1.2  
>>/home/morama/gas/pili/align/orf84.pep (696 aa)  
initn: 83 init1: 57 opt: 135 Z-score: 75.1 expect(): 1.2  
Smith-Waterman score: 146; 24.4% identity in 262 aa overlap  
(232-462:316-567)

```

      210      220      230      240      250
gi-50913506. KTTSLGSLVSLTFELPSYTKAEVNTVYVSDNMSEGLTFNFNSLTVEWKGKMAN----
               |::: |::: :: |::: :
orf84.pep     EKEVAVDDAELKQINSEGQQEISVTWTNQLVTDE--KGMAYIYSVKEVDKNGELLEPKDY
               290      300      310      320      330      340

      260      270      280      290      300      310
gi-50913506. ITEDGSVMVENTKIGIAKEVNNGFNLSFIY----DSLESISPNISYKAVVNKAIIVGEE
               |::: :: |::: |::: :: |::: :: |::: :: |::: :: |::: :: |::: ::
orf84.pep     IKKEDGLTNTYV---KPTSGHYDIEVTFGNHIDITEDTTPDI-VSGENQMKQIEGED
               350      360      370      380      390

      320      330      340      350
gi-50913506. GNP-----NKAFFYSNNP-----TKGNTYDNL-DKKP-DKNGGITSKEDSKIIVTYQ
               ::| |::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |:::
orf84.pep     SKPIDEVTENNLIIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLGSLSEQQGSGDMTIE
               400      410      420      430      440      450

      360      370      380      390      400      410
gi-50913506. -----IAFRKVDVSKTPLIGAIFGVYDTSNKLIDIVTTNKNGYAISTQVSSGKYIKE
               |||| :: ||:: |::: |::: |::: |::: |::: |::: |::: |::: |::: |:::
orf84.pep     EDSATHIKFSKRD-IDGKELAGATMELRDSSGKTIS--TWISDGQVKDFYLMPGKYTFVE
               460      470      480      490      500      510

      420      430      440      450      460      470
gi-50913506. LKAPKGYSLNTE-TYEITAN-WVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGIF
               |||| :: ||:: |::: |::: |::: |::: |::: |::: |::: |::: |::: |:::
orf84.pep     TAAPDGYEIATAITFTVNEQQGVTVNGKATKGDAHVMV-DAYKPTKSGQGVIEDIEEKLP
               520      530      540      550      560      570

      480      490      500      510      520      530
gi-50913506. YSIDSRPITGNDVKEAYTESTKALTDGTTFSKSNESGTFVLETDIPNTKLGELPSTGSIG
               |::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |:::
orf84.pep     DEQCHSGSTTEIEDSKSSDVLIGGQGGIVETTEDTQGMHGDGCKTEVEDTKLVQSFHF
               580      590      600      610      620      630

```

gi-50913506.pep  
/home/morama/gas/pili/align/gi-19745307.pep

gi|19745307|ref|NP\_606443.1| protein F2-like protein [Streptococcus pyogenes MGA S8232]

SCORES Init1: 57 Initn: 83 Opt: 135 z-score: 75.1 E(): 1.2  
>>/home/morama/gas/pili/align/gi-19745307.pep (696 aa)  
initn: 83 init1: 57 opt: 135 Z-score: 75.1 expect(): 1.2  
Smith-Waterman score: 146; 24.4% identity in 262 aa overlap  
(232-462:316-567)

FIGURE 58C

```

      210      220      230      240      250
gi-50913506. KTTSLGSLVLSYSLTFELPSYTKAEVNTVYVSDNMSEGLTFNFNSLTVEWKGMAN----
      |::: :|::: :|::: :|::: :|::: :
gi-19745307. EKEVAVDDAELKQINSEGGQEISVTWTNQLVTDE--KGMAYIYSVKEVDKNGELLEPKDY
      290      300      310      320      330      340

      260      270      280      290      300      310
gi-50913506. ITEDGSVMVENTKIGIAKEVNNGFNLSFIY-----DSLESISPNIISYKAVVNNKAIVGEE
      |::: :|::: :|::: :|::: :|::: :|::: :
gi-19745307. IKKEDGLTVTNTYV---KPTSGHYDIEVTFGNHIDITEDTTPDI-VSGENQMKQIEGED
      350      360      370      380      390

      320      330      340      350
gi-50913506. GNP-----NKAEFFYSNNP-----TKGNTYDNL-DKKP-DKNGGITSKEDSKIVYTYQ
      ::| |::: :|::: :|::: :|::: :|::: :|::: :
gi-19745307. SKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLGLSSEQQQSGDMTIE
      400      410      420      430      440      450

      360      370      380      390      400      410
gi-50913506. -----IAFRKVDVSKTFLIGAIFGVYDTSNKLIDIVITNKNKYAISTQVSSGKYKIKE
      | | | |::: :| | | |::: :| | | |::: :| | | |::: :| | | |::: :
gi-19745307. EDSATHIKFSKRD-IDGKELAGATMELRDSSGKTIS--TWISDGQVKDFYLMFGKYTFVE
      460      470      480      490      500      510

      420      430      440      450      460      470
gi-50913506. LKAPKGYSLNTE-TYEITAN-WVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGIF
      | | | |::: :| | | |::: :| | | |::: :| | | |::: :| | | |::: :
gi-19745307. TAAPDGYELATAITFTVNEQGQVTVNGKATKGDHIVMV-DAYKPTKSGQVIDIEEKLP
      520      530      540      550      560      570

      480      490      500      510      520      530
gi-50913506. YSIDSRPTGNDVKRAYIESTKALTDGTTFSKSNESGTVLLETDPNTKLGLPSTGSIG
gi-19745307. DEQGHSGSTTEIEDSKSSDVIIGGQGVIVETTEDTQTGMHGDGCKTEVEDTKLVQSFHF
      580      590      600      610      620      630

```

gi-50913506.pep

/home/morana/gas/pili/align/gi-21909640.pep

gi|21909640|ref|NP\_663908.1| protein F2 like fibronectin-binding protein [Streptococcus pyogenes MGAS315]

SCORES: Init1: 56 Initn: 81 Opt: 134 z-score: 75.0 E(): 1.2  
 >>/home/morana/gas/pili/align/gi-21909640.pep (656 aa)  
 initn: 81 init1: 56 opt: 134 z-score: 75.0 expect(): 1.2  
 Smith-Waterman score: 156; 23.9% identity in 347 aa overlap  
 (232-547:202-524)

```

      210      220      230      240      250
gi-50913506. KTTSLGSLVLSYSLTFELPSYTKAEVNTVYVSDNMSEGLTFNFNSLTVEWKGMAN----
      |::: :|::: :|::: :|::: :|::: :
gi-21909640. EKEVAVDDAELKQINSEGGQEISVTWTNQLVTDE--KGMAYIYSVKEVDKNGELLEPKDY
      180      190      200      210      220

      260      270      280      290      300      310
gi-50913506. ITEDGSVMVENTKIGIAKEVNNGFNLSFIY-----DSLESISPNIISYKAVVNNKAIVGEE
      |::: :|::: :|::: :|::: :|::: :|::: :
gi-21909640. IKKEDGLTVTNTYV---KPTSGHYDIEVTFGNHIDITEDTTPDI-VSGENQMKQIEGED
      230      240      250      260      270      280

```

FIGURE 58D

PCT/US05/27239

```

          320          330          340          350
gi-50913506. GNP-----NKAEFFYSNNP-----TKGNTYDNL-DKKP-DKNGGITSKEDSKIVYTYQ
          ::|          || ::|          ||| ::|          ||| ::|          ||| ::|
gi-21909640. SKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTL SGLSSEQGQSGDMTIE
          290          300          310          320          330          340

          360          370          380          390          400          410
gi-50913506. -----IAFRKVDVSKTPLIGAIFGVYDTSNKLIDIVT'NKNGYAISTQVSSGKYKIKE
          ||| || ::|          ||| ::|          ||| ::|          ||| ::|
gi-21909640. EDSATHIKFSKRD-IDGKELAGATMELRDSGKTIS--TWISDGQVKDFYLMPGKYTFVE
          350          360          370          380          390          400

          420          430          440          450          460          470
gi-50913506. LKAPKGYSLNTE-TYEITAN-WVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGIF
          || || : | | : : : || :: || : : : : : || : | | : ||
gi-21909640. TAAPDGYEVATAITFTVNEQGQVTVNGKATKGDHIVMV-DAYKPTKGSQGV-----
          410          420          430          440          450

          480          490          500          510          520          530
gi-50913506. YSIDSRPTGNDVKEAYIESTKALTDTGTTFSKSNESGSGTVLLETDPNTKLGEPLSTGSIG
          || : | : : : || : | : : : | : | : | : | : | : |
gi-21909640. -IDIEEKLPD-EQGHSGSTTEIEDSKSSDVIIGGQGEVVDTE--DTQSGMTGHSGST-
          460          470          480          490          500

          540          550
gi-50913506. TYLFKAIGSAAMIGAIGIYIVKRRKA
          | : : : | : : | : |
gi-21909640. TEIEDSKSSDVIIGGQGEVVDTEEDTQSGMTGHSGSTTKIEDSKSSDVIVGGQGQIVETT
          510          520          530          540          550          560

```

gi-50913506.pep

/home/morama/gas/pili/align/gi-28810263.pep

gi|28810263|dbj|BAC63201.1| protein F2-like protein [Streptococcus pyogenes SSI-1]

SCORES Init1: 56 Initn: 82 Opt: 134 z-score: 74.7 E(): 1.3

&gt;&gt;/home/morama/gas/pili/align/gi-28810263.pep (733 aa)

initn: 82 init1: 56 opt: 134 z-score: 74.7 expect(): 1.3

Smith-Waterman score: 155; 23.6% identity in 347 aa overlap  
(232-547:316-638)

```

          210          220          230          240          250
gi-50913506. KTTSLGSLVLSYSLTTELPSTYTKFAVNTVYVSDNMSEGLTNNENSLTVEWKGKMAN-----
          ||| : | : | : | : | : | : | : | : | : | : | : |
gi-28810263. EKEVAVDDAEIKQINSEGGQGEISVTWNLQVTEDE--KGMAYTYSVKEVDKNGELLEPKDY
          290          300          310          320          330          340

          260          270          280          290          300          310
gi-50913506. ITEDGSVMVENTKIGIAKEVNNGFNLSFIY-----DSLSEISPNISYKAVVNNKAIVGEE
          | : : : | || : | : : : : | : : | : | : : | : |
gi-28810263. IKKEDGLTVNTFYV--KPTSGHYDIEVTFNGHIDITEDTTPDI-VSGENQMKQIEGED-
          350          360          370          380          390

          320          330          340          350
gi-50913506. GNP-----NKAEFFYSNNP-----TKGNTYDNL-DKKP-DKNGGITSKEDSKIVYTYQ
          ::|          || ::|          ||| ::|          ||| ::|          ||| ::|
gi-28810263. SKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTL SGLSSEQGQSGDMTIE
          400          410          420          430          440          450

          360          370          380          390          400          410

```

FIGURE 58E

PCT/US2005/027239

```

gi-50913506. -----IAFRKVDVSVKTPPLIGAFGVYDTSNKLIDIVTINKNGYAISTQVSSGKYKIKE
| | | | :: | | : : | : | | : | : : : : | | : |
gi-28810263. EDSATHIKFSKRD-IDGKELAGATMELRDSSGKTIS--TWISDGOVKDFYLMPGKYTFVE
460      470      480      490      500      510

      420      430      440      450      460      470
gi-50913506. LKAPKGYSINTE-TYEITAN-WVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGIF
| | | : | | : : : | | : | : : : : : | | | : | |
gi-28810263. TAAPDGYEVATAITFTVNEQGQVTVNGKATKGAHIVMV-DAYKPTKSGSQV-----
520      530      540      550      560

      480      490      500      510      520      530
gi-50913506. YSIDSRPTGNDVKEAYIESTKALTDGTTFSKSNEGSGTVLLETDPINTKLGLPSTGSIG
| | : | : : : | | : | : : : | : | : | : | : |
gi-28810263. --IDIEKLPD-EQGHSGSTTEIEDSKSSDVIIGGQGEVVDTE--DTQSGMTGHSGST-
570      580      590      600      610      620

      540      550
gi-50913506. TYLFKAIGSAAMIGAIGIYIVKRRKA
| : : : | : : : |
gi-28810263. TKIEDSKSSDVIVGGQGIIVETTEDTQTMHGDSGRKTEVEDTKLVQSFHFDNKEPESNS
630      640      650      660      670      680

```

gi-50913506.pep  
/home/morama/gas/pili/align/orf80.pep

TRANSLATE of: orf80.seq check: 9824 from: 1 to: 1056  
generated symbols 1 to: 352.  
GETSEQ from morama, September 13, 2004 17:11.

SCORES Init1: 45 Initn: 69 Opt: 113 z-score: 70.8 E(): 2.1  
>>/home/morama/gas/pili/align/orf80.pep (352 aa)  
initn: 69 init1: 45 opt: 113 Z-score: 70.8 expect(): 2.1  
Smith-Waterman score: 123; 22.8% identity in 311 aa overlap  
(284-556:49-352)

```

      260      270      280      290      300      310
gi-50913506. KMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLESISPNI SYKAVVN-NKAIVGEE
| | | | : | : | : : : : : | |
orf80.pep. ATALGTASLNQNVKAETAGVVTGKSLQVTKMTYDDEEVLMPEAFTFTTIEPDMTASGKE
20      30      40      50      60      70

      320      330      340      350      360      370
gi-50913506. GNPN-KAEFFYSNNPTKGNITYDNLDRKPKDKNCGITSKEDSKIVYTYQIAFRKVDVSVKTP
| : : | : : : | | | : | : : | : : : | : :
orf80.pep. GSLDIKNGIVEGLDKQVTVKYNTRPSQTK-LAQDFSKVKFPAIGVYRYMVSEKNDK
80      90      100      110      120      130

      380      390      400      410      420
gi-50913506. LIGAFGVYDTSNKLIDIVTINKNG-----YAISTQ-VSSGKYKIKELKAPKGYSIN
| : | : : : | : : | : : : | : : : | : :
orf80.pep. KDGI---TYDDKKWTVDVYVGKANNNEEGFEVLYIVSKEGTSSTKKPIEFTNSIKTTSLK
140      150      160      170      180      190

      430      440      450      460
gi-50913506. TETYEITANW-----VTATVKTSANSKSTTYTSDKNKATDNSEQVG-----WLKNGI
| : | : | : | : | : : : : : : : : :
orf80.pep. IEK-QITGNAGDRKKSFNFTLTLPSEYYKTGSVVKIEQDGSKKDVTIGTPYKFTLGHGK
200      210      220      230      240      250

470      480      490      500      510

```

FIGURE 58F

PCT/US05/27239

gi-50913506. FYSIDSRPTGNDV-----KEAYI-----ESTKALTDGTTFSKSNESGTVLLETDI  
 :: | | : |:: | | : | | : | | : | | :  
 orf80.pep SVMLSKLPIGINYLLSEDEANKDGYTTTATLKEQGKEKSSDFTLSTQNKTKDESADIVV  
 260 270 280 290 300 310

520 530 540 550  
 gi-50913506. PNTKLGEPLSTGSIGTYLFKAIGSAAMIGAIGIYIVKRRKA  
 | : :: | | : | | : | | : | | : | | : | | :  
 orf80.pep TNKRDTQVP-TGVVGTLPAPFAVL SIVAIGGV-IYITKRRKA  
 320 330 340 350

gi-50913506.pep  
 /home/morاما/gas/pili/align/gi-19224137.pep

gi|19224137|gb|AAL86408.1|AF447492\_5 EftLSL.A [Streptococcus pyogenes]

SCORES Initl: 45 Initn: 69 Opt: 109 z-score: 69.8 E(): 2.4  
 >>/home/morاما/gas/pili/align/gi-19224137.pep (342 aa)  
 initn: 69 initl: 45 opt: 109 z-score: 69.8 expect(): 2.4  
 Smith-Waterman score: 169; 26.0% identity in 334 aa overlap  
 (257-556:25-342)

230 240 250 260 270 280  
 gi-50913506. NKTVYVSDNMSEGLTFNFNLSLVEWKGKMANITEDGSVMVENTKIGIAKEVNNGFNLSFI  
 | : : : | : : : | | : | |  
 gi-19224137. MKKNKLLLATAILATALGTASLNQNVKAETAGVVSSGQLTIKKSIITN-FN----  
 10 20 30 40

290 300 310 320 330 340  
 gi-50913506. YDSLESISPNI SYKAVVN-NKAIVGEEGN-PNKAIEFFYSNNPTKGNTYDNLDDKPKDKNNG  
 | : | : | : | | : | : | : | : | : | : | : | :  
 gi-19224137. DDTL--LMPKTDYTF SVNPD SAATGTESNLPIKPGIAVNNQDIK-VSYNNTDKTSGKEKQ  
 50 60 70 80 90 100

350 360 370 380 390 400  
 gi-50913506. ITSKEDESKIVYTYQIAFRKVDVSVKTIPLIGAIFGV-YDTSNKLIDIVTNNKNGYAISTQV  
 : : : | : : : | : | : | : | : | : : : : : : : : :  
 gi-19224137. VVV-DFMKVTFPSVGIYRYVVTENK---GTAEGVTYDDTKWLVDVYVGNNEKGGLEPKY  
 110 120 130 140 150 160

410 420 430 440 450  
 gi-50913506. SSGKYKIKELKAPKGY--SLNTETYBITANWVTATVKTSA NSKSTTYTSDKNKATDNS--  
 : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
 gi-19224137. IVSKKGD SATKEPIQFNNSFTTSLKIEKE-VTCNTGDKKAF TTTLTLPNEYEASSV  
 170 180 190 200 210 220

460 470 480 490  
 gi-50913506. ---EQVGWLKN---GIFYSI---DSR-----PTGND--VKEAYIE-----STKALTDG  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
 gi-19224137. VKIEENGQTRDVKIGI EAYKFTLNDSQS VILSKLPVGIN YKVEEAEANQGGYTTTATLKDG  
 230 240 250 260 270 280

500 510 520 530 540 550  
 gi-50913506. TTFSKSNESG---SGTVLLETDI PNTKLGEPLSTGSIGTYLFKAIGSAAMIGAIGIYIVK  
 : | | : : | : : : : | : | : | : | : | : | : | : | : | : | :  
 gi-19224137. EKLSTYNLGQEHKTDKTADEIVVTNNRDTQVP-TGVVGTLPAPFAVL SIVAIGGV-IYITK  
 290 300 310 320 330

gi-50913506. RRKA  
 | : |

FIGURE 58G



PCT/US05/27239

194/487

gi-19224137. RKKA  
340

gi-50913506.pep

/home/morama/gas/pili/align/gi-19224141.pep

gi|19224141|gb|AAL86412.1|AF447492\_9 protein F2 [Streptococcus pyogenes]

SCORES Init1: 73 Initn: 73 Opt: 118 z-score: 68.9 E(): 2.7  
 >>/home/morama/gas/pili/align/gi-19224141.pep (1161 aa)  
 initn: 73 init1: 73 opt: 118 z-score: 68.9 expect(): 2.7  
 Smith-Waterman score: 174; 23.9% identity in 406 aa overlap  
 (115-483:277-645)

```

          90      100      110      120      130      140
gi-50913506. VSLTETGPTSQEITTIANGINTGKIKPFSTENVISISNGTATYNARGASVY--IALLTGAT
          ||:||||| :| ::: : : :| |
gi-19224141. IYTFTDYIAGLDKVLQSAELSLFLENKEVLENTSISNFKSTIGGQEITYKGTNVNVLGYNE
          250      260      270      280      290      300

          150      160      170      180      190
gi-50913506. DGRTYNPILLAASYNAGEGNLVTKNIDSKSNLYLGQTSVAKSSLPSITKKVVG-----T
          : : | | : | | : : | : : | : : : : | | : : |
gi-19224141. STKESNYITNGLSNVG-GSIESYNTETGEFVWVYVNPRTNIPYATMNLWGFGRARSNT
          310      320      330      340      350      360

          200      210      220      230      240      250
gi-50913506. ID---DVNKKTTSLGSVLSYSITF--ELPSYTKAEVNKTVVSDNMSEGLTFNFNLSLTVE
          | | :| : : | | : : : ||| | :| : : | : | :| :| :|
gi-19224141. SDLENDANTSSAELGEIQVYEVPEGEKLPSSYGVDTVTKLTLRTD-ITAGLNGGFQ-----
          370      380      390      400      410

          260      270      280      290      300      310
gi-50913506. WKGKMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLESISPNI SYKAVVNNKAIVG
          | : :| : : || | :|| : : : | | :| : : : :
gi-19224141. -----MTKRQRIDFG---NNIQNKAFIIKV--TGKTDQSGKPLVVQSNLAS
          420      430      440      450      460

          320      330      340      350      360
gi-50913506. EEGNPNKAEFFYSNNPTKGNTRY--DNLDDKPKDKNGITSKEDSKIVYTY-----QIAF
          :| :| | :| :||| : : : :| :||| : :| :| | :| :|
gi-19224141. FRGASEYAAF-----TPVGGNVYFQNEIALSPSKGSGSGKSEETKPSITVANLKRVAQLRF
          470      480      490      500      510

          370      380      390      400      410      420
gi-50913506. RKVDSVSKTPLIGATFGVYDTSNKLIDI-VTTNKNGYATSTQVSSGKYKIKELKAPKGYG
          :| :| : : :| | | : : : : : : :| :| : :| :| :| :| :| :| :| :| :| :| :|
gi-19224141. KKM-STDNVPLPEAFELRSSNGNSQKLEASSNTQGEVHFKDLTSGTYDLYETKAPKGYQ
          520      530      540      550      560      570

          430      440      450      460
gi-50913506. -----LNTETYEIT-----ANWVT--ATVKTSANSKSTTYTSDKNKATDNSEQVGWLKN
          | | | : | : :| : : :||| | :| : :| :| :| : : :| : :
gi-19224141. QVTEKLATVTVDTKPAEMVTWGSPPHSSVKVEAN-KEVTIVNHKETLTFSGKKI-WEND
          580      590      600      610      620      630

          470      480      490      500      510      520
gi-50913506. GIFYSIDSRPTGNDVKEAYIESTKALTDGTTFSKSNESGCTVLLETDPNTKLGELPSTG
          | :|| : :| :
gi-19224141. ----RPDQRFPAKIQVQLLQNGQKMPNQIQEVTKDNDWSYHFKDLPKYDAKNQYKYSVEE
          640      650      660      670      680

```

FIGURE 58H

gi-50913506.pep

/home/morama/gas/pili/align/gi-21909636.pep

gi|21909636|ref|NP\_663904.1| conserved hypothetical protein [Streptococcus pyogenes MGAS315]

SCORES Init1: 45 Initn: 98 Opt: 96 z-score: 66.1 E(): 3.8  
 >>/home/morama/gas/pili/align/gi-21909636.pep (344 aa)  
 initn: 98 init1: 45 opt: 96 Z-score: 66.1 expect(): 3.8  
 Smith-Waterman score: 181; 25.3% identity in 312 aa overlap  
 (298-556:44-344)

```

      270      280      290      300      310      320
gi-50913506. NTKIGIAKEVNNGFNLSFIYDSLESISPNI SYKAVVNNKAIVGEEGNPNKAEFFYSNNPT
      :: : ::|::: : |::| ::: :
gi-21909636. ATALGTASLNQNVKAETAGVSENAKLIVKKTFSYTDNEVLMPKADYTFKVE---ADSTA
      20      30      40      50      60      70

      330      340      350      360
gi-50913506. KGNTYDNLDDKPKDKNGIT-----SKEDSKIVYTYQIAFRKV-----DSVSK
      :|:| |:|: || |:| :| |||: |::| || :||:
gi-21909636. SGKTKDGLEIKPGIVNGLTEQII SYTNTDKPDSKVKST-EFDFSKVVFPGIGVRYTVSE
      80      90      100      110      120

      370      380      390      400      410      420
gi-50913506. TPLIGAIFGV-YDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKELKAPKGY--SINTETY
      | : | : ||:: :|: : ||:| :: : |: : ::| |:: |:|:
gi-21909636. KQ--GDVEGITYDTPKKWTVDVYVGNKEGGGFEPKFIVSKEQGTDVKKPVNFNNSFATTSL
      130      140      150      160      170      180

      430      440      450      460
gi-50913506. EITANWVTATVKT SAN-----SKSTTYTSDK---NKATDNSE-QVGW-----LKNIGI
      :: | | : : : ::||:: :|: :|::: :|::| |||
gi-21909636. KVKKNVSGNTGELQKEFDFTLT LNESTNFKKDQIVSLQKNEKFEVKIGTPYKFKLKNGE
      190      200      210      220      230      240

      470      480      490      500      510
gi-50913506. FYSIDSRPTGNDVKEAYIESTKALTDGTTFSKS-NEGSGTVLL-----ETD-----
      ::|: |:| | :|::| || :| :|::| : :|:
gi-21909636. SIQLDKLPVGITYKVNEMEANK---DGYKTTASLKEGDSKMYQLDMEQKTDESAD EIV
      250      260      270      280      290      300

      520      530      540      550
gi-50913506. IPNTIKLGELPSTG SIGTYLEKAIGSAAMIGAIGIYIVKRRKA
      : | : ::| ||:| | :| :|::| ||:| |::|
gi-21909636. VTNKRDTQVP-TGVVGT LAPFAVL SIVAIGGV-IYITKRRKA
      310      320      330      340

```

gi-50913506.pep

/home/morama/gas/pili/align/gi-28810259.pep

gi|28810259|dbj|BAC63197.1| hypothetical protein [Streptococcus pyogenes SSI-1]

SCORES Init1: 45 Initn: 98 Opt: 96 z-score: 66.0 E(): 3.8  
 >>/home/morama/gas/pili/align/gi-28810259.pep (349 aa)  
 initn: 98 init1: 45 opt: 96 Z-score: 66.0 expect(): 3.8  
 Smith-Waterman score: 181; 25.3% identity in 312 aa overlap  
 (298-556:49-349)

FIGURE 58I

PCT/US05/27239

```

270      280      290      300      310      320
gi-50913506. NTKIGIAKEVNNGFNLSFIYDSLESISPNISYKAVVNNKAIVGEEGNPNKAEFFYSNNPT
                :: : ::|::: : |:: : :
gi-28810259. ATALGTASLNQNVKAETAGVSENAKLIVKKTFSYTDNEVLMPKADYTFKVE---ADSTA
20      30      40      50      60      70

330      340      350      360
gi-50913506. KGNTYDNLDKKPKDKNGIT-----SKEDSKIVYTYQIAFRK-----DSVSK
:|:| |:|: || |:| :| ||: | :: | || :||:
gi-28810259. SGKTKDGLEIKPGIVNGLTEQIISYTNIDKPD SKVKST-EFDFSKVVFPGIGVYRYTVSE
80      90      100      110      120      130

370      380      390      400      410      420
gi-50913506. TPLIGAIFGV-YDTSNKLIDIVITNKNGYAISTQVSSGKYKIKELKAPKGY--SLNTETY
| : |: ||:: :|: : ||:| : : : | : ::| | : : |:| :
gi-28810259. KQ--GDVEGITYDTKKWTVDVYVGNKEGGGFEPKFIVSKEQGTVDVKKPVNFNNSFATTSL
140      150      160      170      180      190

430      440      450      460
gi-50913506. EITANVWTATVKTSAN-----SKSTTYTSDK-----NKATDNSE-QVGW-----LKNGI
: : : | : : : : :|::: :|: :|::: | :| |||
gi-28810259. KVKKNVSGNTGELQKEFDFTLTINESTNFKKDQIVSLQKNEKFEVKIGTPYKFKLKNGE
200      210      220      230      240      250

470      480      490      500      510
gi-50913506. FYSIDSRPTGNDVKEAYLESTKALTDGTTESKS-NEGSGTVLL-----ETD-----
:|:| |:| | :|:| || :| :|::| : :|
gi-28810259. SIQLDKLPVGITYKVNEMEANK---DGYKTASLKEGDGQSKMYOLDMEQKTDSEADEIV
260      270      280      290      300

520      530      540      550
gi-50913506. IPNTKLGELPSTGSIGTYLFAIGSAAMIGAIGIYIVKRRKA
: | : ::| ||: || |: | : ||:: ||: ||: ||
gi-28810259. VTNKRDTQVP-TGVVGTLPAPFAVL SIVAIGGV-IYITKRKKA
310      320      330      340

```

gi-50913506.pep  
/home/morana/gas/pili/align/gas15.pep

GAS15 GAS15

SCORES Initl: 42 Initn: 68 Opt: 96 z-score: 63.8 E(): 5  
>>/home/morana/gas/pili/align/gas15.pep (762 aa)  
initn: 68 initl: 42 opt: 96 z-score: 63.8 expect(): 5  
Smith-Waterman score: 96; 23.4% identity in 269 aa overlap  
(283-535:222-470)

```

260      270      280      290      300
gi-50913506. GKMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLESISPNISYKAVVNN--K
                ||:: ::|::: |:|::: | :|
gas15.pep    VWYYS DNAPISNPDESFKRESESNLVSTSLSLMRQALKQLIDPNLATKMPKQVPDDFQL
200      210      220      230      240      250

310      320      330      340      350      360
gi-50913506. AIVGEEGNPNKAEFFYSNNPTKGNTYDNLDKKPKDKNGIT'SKEDSKIVYTYQIAFRK--V
:| | : |:| :|: | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
gas15.pep    SIFESDKGDKYNKGYQNLLSGGLVPT---KPPTPGDPMPMPNQPQ---TTSVLIRKYAI
260      270      280      290      300

370      380      390      400      410      420

```

FIGURE 58J

**FIGURE 58K**

FIGURE 58L



199/487

gi-19224135. YNKKGDSYNGLTETQFRAATQLAIYYFTDSTDCLKTLKTYNNGKGYHGFESMDEKTLAVTK  
 490 500 510 520 530 540

gi-50913506.pep

/home/morana/gas/pili/align/gi-50913505.pep

gi|50913505|ref|YP\_059477.1| Collagen adhesion protein [Streptococcus pyogenes M  
 GAS10394]

SCORES Init1: 70 Initn: 149 Opt: 86 z-score: 60.1 E(): 7.8  
 >>/home/morana/gas/pili/align/gi-50913505.pep (1036 aa)  
 .initn: 149 init1: 70 opt: 86 z-score: 60.1 expect(): 7.8  
 .Smith-Waterman score: 120; 21.5% identity in 469 aa overlap  
 (33-428:503-966)

```

      10      20      30      40      50      60
gi-50913506. NRRET VREKILITAKKLM LACLAILAVVGLG MTRVS-ALSKDDTAQLKITNIEGGPTVTL
      |:: || :: :||::: :: ::::|
gi-50913505. YSQWVYHGTFDYSSYQGESFNRGQIHRYRGIVSVSDGIRRDDAVKNSLLGVNG-----L
      480      490      500      510      520

      70      80      90      100     110
gi-50913506. YKIGEGVYNTNGDSFINFK----YAEGVSLTETGPTSQBIT-TIANGINTGKIKPFSTEN
      : :: | | :|: | | : :|: : :: |:::
gi-50913505. LQRFVNINPENKLSVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNSRDAELLKGWSTNS
      530      540      550      560      570      580

      120     130     140     150     160
gi-50913506. VVISNGTATYNARGASVYIALITGAT-----DGRTYNPILLAAS-----YNGEGNLTVK
      : | : : : :|: : ||| : ||| :|: : | || : :
gi-50913505. LLDENTLTALHNNGTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPTFFYFGEDGYRSG
      590      600      610      620      630      640

      170     180     190     200     210
gi-50913506. NIDS--KSNLYLGQ--TSVA---KSSLPSITKKVTGTIDDVNKKTTSLGSVLSYSLTFE
      | :| :|: :| :|: :|: :|: :|: :|: :|: :|:
gi-50913505. NGSSNDRNNVTRSQEGSKLAIDEFKARYPNLSIYSLGVSKDINSDTASSPVVLKYLKSGE
      650      660      670      680      690      700

      220     230     240     250     260
gi-50913506. LPSY-----TKEAVNKIVY-----VSDNMSEGLTF--NFNSLTVEWKGKMANITE
      | :|: :|: :|: :|: :|: :|: :|: :|: :|:
gi-50913505. EHYYGITDTAELEKTNKIVEDSKLSQLGISDLSQYVDYDKQPDVLVTRKSKVNDETE
      710      720      730      740      750      760

      270     280     290
gi-50913506. -----DGSVMVENTKIGIAKEVNNGFNLSFIYDS--LESISPNI-----
      | :|: :|: :|: :|: :|: :|: :|: :|:
gi-50913505. ILYQKDQVQEAGKDIIDKVVTTPKTTSQPKGVTLTFKSDYKVDDEYTYTLTFNVKASDE
      770      780      790      800      810      820

      300     310     320     330     340     350
gi-50913506. SYKAVVNNKAI VGEENFENKAEFFYSNNPTKGNTYDNLDDKPKDK-GNGITSKEDSK----
      :|: :|: :|: :|: :|: :|: :|: :|: :|:
gi-50913505. AYEKYDNEGRYSEMGDSDTDYGTNQTSSGKGGLPSNSDASVNYMADGREQKLPYKHPVI
      830      840      850      860      870      880

      360     370     380     390     400
gi-50913506. IVYTYQIAFRKVDS-VSKTPLIGAIFGVYDTSNKLI-DIVTTNKNGYAISTQVSSGK-YK
      | | :| |||: : :| :| : :|: :| :|: :| :|:

```

FIGURE 58M

**FIGURE 58N**

gi-19745303. KGRKTKDGLDLPKPGVLDGIBENTKTIHYGNSDKTTAKEKSVNFDANVKFPGVGVYRYTVSE  
80 90 100 110 120 130

gi-50913506. LIGAIFGV-YDTSNKLIDIVTTNKN--GYAISTQVSS-GKYKIKELKAPKGYSLNTETYE  
380 390 400 410 420  
: | |::: |: :::: |: : ||: | : |: |:::| : :  
gi-19745303. VNGNKAGIAYDSQQWTVVDVYVVNREDGGFEAKYIVSTEGGQSDKKPVLFKNF-FDITSLK  
140 150 160 170 180

gi-50913506. ITANWVTATVKTSAKSTTYYTSDKNKATDNSEQVGWLKNGIF-----YSIDSR--  
430 440 450 460 470  
: | : ||::: : | | | |: :::: |: |::: |::: :  
gi-19745303. VTKK-VTGNTEGEHQRSFSFTLLLTNPNECFEKGQVNNILQGGETKKVVIGREYSFTLKDKKE  
190 200 210 220 230 240

gi-50913506. -----PTG-----NDVKEAYIESTKALTDTTFSSKSNEG-SGTVLLETD--IPNT  
480 490 500 510  
|:| :|| : :::: | || : : | | | |: || : |  
gi-19745303. SVTLSQLPVGIEYKVTEEDVTKEGKYKTSATLKDGVDTCYNLGDGSKTTDKSTDEIVVTNK  
250 260 270 280 290 300

gi-50913506. KLGLPSTGSGTYLFAIGSAAMIGAIGIYIVKRRKA  
520 530 540 550  
: :| || :|| |: | : ||:: ||::||:|  
gi-19745303. RDTQVP-TGVVGTLPAPFAVLISIVAIGGV-IYITKRKA  
310 320 330 340

! Distributed over 1 thread.

! Start time: Wed Sep. 15 18:45:54 2004

! Completion time: Wed Sep. 15 18:46:02 2004

! CPU time used:

! Database scan: 0:00:00.1

! Post-scan processing: 0:00:01.9

! Total CPU time: 0:00:02.0

! Output File: gi-50913506.fasta

FIGURE 580

!!SEQUENCE LIST 1.0

(Peptide) FASTA of: gi-13621430.pep from: 1 to: 215 September 15, 2004 18:45

gi|13621430|gb|AAK33240.1| hypothetical protein [Streptococcus pyogenes]

TO: \*.pep Sequences: 56 Symbols: 22,803 Word Size: 2

Scoring matrix: GenRunData:blosum50.cmp

Variable pamfactor used

Gap creation penalty: 12 Gap extension penalty: 2

## Histogram Key:

Each histogram symbol represents 1 search set sequences:

z-scores computed from opt scores

z-score	obs	exp
(=)	(*)	
< 20	0	0:
22	0	0:
24	0	0:
26	0	0:
28	0	0:
30	0	0:
32	0	0:
34	0	1:*
36	0	2: *
38	0	3: *
40	0	4: *
42	0	5: *
44	1	5:= *
46	7	5:====*==
48	21	5:====*=====
50	7	5:====*==
52	2	4:== *
54	4	3:==*==
56	3	3:==*
58	0	2: *
60	4	2:==*==
62	1	2:==*
64	0	1:*
66	0	1:*
68	0	1:*
70	0	1:*
72	0	0:
74	0	0:
76	0	0:
78	0	0:
80	3	0:===
82	2	0:==
84	0	0:
86	0	0:
88	0	0:
90	0	0:
92	0	0:
94	0	0:
96	0	0:
98	0	0:
100	0	0:

FIGURE 59

```

102 PCT/US2005/027239
104 0 0:
106 0 0:
108 0 0:
110 0 0:
112 0 0:
114 0 0:
116 0 0:
118 0 0:
>120 1 0:=

```

Joining threshold: 36, opt. threshold: 24, opt. width: 16, reg.-scaled

The best scores are:

	init1	initn	opt	z-sc	E(55)
/home/morama/gas/pili/align/gi-13621430.pep	Begin: 1	End: 215			
! gi 13621430 gb AAK33240.1  hypothet... 1338	1338	1338	233.9	1.8e-09	
/home/morama/gas/pili/align/gi-19745305.pep	Begin: 1	End: 193			
! gi 19745305 ref NP_606441.1  hypoth... 163	243	273	82.2	0.5	
/home/morama/gas/pili/align/gi-28810261.pep	Begin: 2	End: 187			
! gi 28810261 dbj BAC63199.1  hypothe... 164	239	268	81.5	0.55	
/home/morama/gas/pili/align/gi-19224139.pep	Begin: 2	End: 187			
! gi 19224139 gb AAL86410.1 AF447492... 164	236	265	81.0	0.57	
/home/morama/gas/pili/align/orf82.pep	Begin: 30	End: 222			
! TRANSLATE of: orf82.seq check: 4296... 163	235	264	81.0	0.58	
/home/morama/gas/pili/align/gi-21909638.pep	Begin: 2	End: 181			
! gi 21909638 ref NP_663906.1  hypoth... 164	239	261	80.5	0.62	
/home/morama/gas/pili/align/gi-19745303.pep	Begin: 84	End: 183			
! gi 19745303 ref NP_606439.1  hypoth... 121	121	126	61.4	6.7	
/home/morama/gas/pili/align/gi-13621428.pep	Begin: 6	End: 174			
! gi 13621428 gb AAK33238.1  hypothet... 58	86	122	60.9	7.2	
/home/morama/gas/pili/align/gi-19224137.pep	Begin: 93	End: 201			
! gi 19224137 gb AAL86408.1 AF447492... 88	88	119	60.4	7.5	
/home/morama/gas/pili/align/gi-50913503.pep	Begin: 549	End: 625			
! gi 50913503 ref YP_059475.1  Fibron... 73	73	117	60.4	7.6	
/home/morama/gas/pili/align/gi-19224134.pep	Begin: 631	End: 697			
! gi 19224134 gb AAL86405.1 AF447492... 73	73	115	60.1	7.8	

\\End of List

gi-13621430.pep

/home/morama/gas/pili/align/gi-13621430.pep

gi|13621430|gb|AAK33240.1| hypothetical protein [Streptococcus pyogenes]

SCORES Init1: 1338 Initn: 1338 Opt: 1338 z-score: 233.9 E(): 1.8e-09  
 >>/home/morama/gas/pili/align/gi-13621430.pep (215 aa)  
 initn: 1338 init1: 1338 opt: 1338 z-score: 233.9 expect(): 1.8e-09  
 Smith-Waterman score: 1338; 100.0% identity in 215 aa overlap  
 (1-215:1-215).

```

      10      20      30      40      50      60
gi-13621430. MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
              |||
gi-13621430. MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
              |||
      10      20      30      40      50      60

      70      80      90     100     110     120
gi-13621430. EALDKESPLPNSVTTSVKNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVYEVV
              |||
gi-13621430. EALDKESPLPNSVTTSVKNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVYEVV
              |||

```

FIGURE 59A

PCT/US2005/027239

```

          90          100          110          120
    130      140      150      160      170      180
gi-13621430. IYVLYNEQSGALETNLVSNKLGETEKSSELIFKQEYSEKTPPEHPQDPTTEKEKPQKKRNGI
|||||
gi-13621430. IYVLYNEQSGALETNLVSNKLGETEKSSELIFKQEYSEKTPPEHPQDPTTEKEKPQKKRNGI
          130          140          150          160          170          180
          190          200          210
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
|||||
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
          190          200          210

```

gi-13621430.pep

/home/morama/gas/pili/align/gi-19745305.pep

gi|19745305|ref|NP\_606441.1| hypothetical protein [Streptococcus pyogenes MGAS8232]

SCORES Init1: 163 Initn: 243 Opt: 273 z-score: 82.2 E(): 0.5  
 >>/home/morama/gas/pili/align/gi-19745305.pep (195 aa)  
 initn: 243 init1: 163 opt: 273 z-score: 82.2 expect(): 0.5  
 Smith-Waterman score: 320; 31.9% identity in 213 aa overlap  
 (1-213:1-193)

```

          10          20          30          40          50          60
gi-13621430. MKKSILRLAIGYLLMSFCLLDSEVAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
|:| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
gi-19745305. MRKYWKMLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTS-----FSVAL
          10          20          30          40          50
          70          80          90          100          110          120
gi-13621430. EALDKESPLPNSVTTTSVKGNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVEYEVV
|:| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
gi-19745305. ESIDAMKTI-DEIT--IAGSGKASFSPLETFVVGQYTYRVYQKPSQNKDYQADTTVFDVL
          60          70          80          90          100          110
          130          140          150          160          170          180
gi-13621430. IYVLYNEQSGALETNLVSNKLGETEKSSELIFKQEYSEKTPPEHPQDPTTEKEKPQKKRNGI
|:| |:| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
gi-19745305. VYVYDDED-GTLVAKVISRRAGDEEKSATTFKPKRLVKPIPPRPQDI-----PKTF-----
          120          130          140          150          160
          190          200          210
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
|:| |:| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
gi-19745305. LPLAGEVKSLILGILSIVLLGLLVLLYV-KKIKSRL
          170          180          190

```

gi-13621430.pep

/home/morama/gas/pili/align/gi-28810261.pep

gi|28810261|dbj|BAC63199.1| hypothetical protein [Streptococcus pyogenes SSI-1]

SCORES Init1: 164 Initn: 239 Opt: 268 z-score: 81.5 E(): 0.55  
 >>/home/morama/gas/pili/align/gi-28810261.pep (189 aa)  
 initn: 239 init1: 164 opt: 268 z-score: 81.5 expect(): 0.55  
 Smith-Waterman score: 306; 30.6% identity in 206 aa overlap

FIGURE 59B



(8-2132-187) US 05/27239

```

      10      20      30      40      50      60
gi-13621430. MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
      :::: ::| : : ::| |:: |:: :| | : : | : :
gi-28810261.  MLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTP-----FSIAL
      10      20      30      40

      70      80      90      100     110     120
gi-13621430. EALDKESPLPNSVTTTSVKNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVEYEVV
      |::| : : : : : : |::| |::| |::| |::| |::| |::| |::|
gi-28810261. ESIDAMKTIEE---ITTAGSGKASESPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFVDVL
      50      60      70      80      90      100

      130     140     150     160     170     180
gi-13621430. IYVLYNEQSGALETNLVSNNKLGETEKSELIFKQEYSEKTPEPHQPDTEKEKPKQKKRNGI
      :|| |::| :| :::| : |::| : || : : | :||: |
gi-28810261. VYVTYDED-GTLVAKVISRRAGDEEKSAITFKPKWLVPPIPPROPNI-----PKTP----
      110     120     130     140     150

      190     200     210
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
      || :||: | : : |::|:: : :| : |||:
gi-28810261. LPLAGEVKSLGLSIVLLGLLVLLYV-KKLSKRL
      160     170     180

```

gi-13621430.pep

/home/morana/gas/pili/align/gi-19224139.pep

gi|19224139|gb|AAL86410.1|AF447492\_7 unknown [Streptococcus pyogenes]

```

SCORES   Init1: 164   Initn: 236   Opt: 265   z-score: 81.0   E(): 0.57
>>/home/morana/gas/pili/align/gi-19224139.pep (189 aa)
initn: 236 init1: 164 opt: 265 Z-score: 81.0 expect(): 0.57
Smith-Waterman score: 303;   30.6% identity in 206 aa overlap
(8-213:2-187)

```

```

      10      20      30      40      50      60
gi-13621430. MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
      :::: ::| : : ::| |:: |:: :| | : : | : :
gi-19224139.  MLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTP-----FSIAL
      10      20      30      40

      70      80      90      100     110     120
gi-13621430. EALDKESPLPNSVTTTSVKNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVEYEVV
      |::| : : : : : : |::| |::| |::| |::| |::| |::|
gi-19224139. ESIDAMKTIEE---ITTAGSGKASESPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFVDVL
      50      60      70      80      90      100

      130     140     150     160     170     180
gi-13621430. IYVLYNEQSGALETNLVSNNKLGETEKSELIFKQEYSEKTPEPHQPDTEKEKPKQKKRNGI
      :|| |::| :| :::| : |::| : || : : | :||: |
gi-19224139. VYVTYDED-GTLVAKVISRRAGDEEKSAITFKPKRLVKPIPPROPNI-----PKTP----
      110     120     130     140     150

      190     200     210
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
      || :||: | : : |::|:: : :| : |||:
gi-19224139. LPLAGEVKSLGLSIVLLGLLVLLYV-KKLSKSL
      160     170     180

```

FIGURE 59C

PCT/US05/27239

gi-13621430.pep  
/home/morama/gas/pili/align/orf82.pep

TRANSLATE of: orf82.seq check: 4296 from: 1 to: 672  
generated symbols 1 to: 224.

GETSEQ from morama, September 13, 2004 17:09.

SCORES Init1: 163 Initn: 235 Opt: 264 z-score: 81.0 E(): 0.58  
>>/home/morama/gas/pili/align/orf82.pep (224 aa)  
initn: 235 init1: 163 opt: 264 Z-score: 81.0 expect(): 0.58  
Smith-Waterman score: 304; 30.5% identity in 213 aa overlap  
(1-213:30-222)

```

gi-13621430.      10      20      30
                  MKKSILRILAIGYLLMSFCLLDSVEAENLTA
orf82.pep      LLFQRVKIFLLTIVLSLSVLFKNNERRRLRKYWKMLFSVVMILTMLAFNQTVLAKDSTV
                  10      20      30      40      50      60
gi-13621430.      40      50      60      70      80      90
                  SINIEVINQVDVATNKQSSDIDETFMFVIEALDKESPLNSVTSVKGNGKTSFEQLTFS
orf82.pep      QTSISVENVLERAGDSTP-----FSVALESIDAMKTI-DEIT--IAGSGKASFSPLTFT
                  70      80      90      100     110
gi-13621430.      100     110     120     130     140     150
                  EVGQYHYKIHQLLGKNSQYHYDETVEVVIYVLYNEQSGALETNLVSNKLGETEKSELIF
orf82.pep      TVGQYTYRVYQKPSQNKDYQADTTVFVDLVVYTYDED-GTLVAKVISRRAGDEEKSAITF
                  120     130     140     150     160     170
gi-13621430.      160     170     180     190     200     210
                  KQEYSEKTPEPHQPDTEKEKPKKRNGLPSTGEMVSYVSALGIVLVATITLYSIYKKL
orf82.pep      KPKRLVKPIPPRPQNI-----PKTP-----LPLAGEVKSLLGILSIVLLGLLVLLVY-KKL
                  180     190     200     210     220
gi-13621430.      KTSK
orf82.pep      KSRL

```

gi-13621430.pep  
/home/morama/gas/pili/align/gi-21909638.pep

gi|21909638|ref|NP\_663906.1| hypothetical protein [Streptococcus pyogenes MGAS31  
5]

SCORES Init1: 164 Initn: 239 Opt: 261 z-score: 80.5 E(): 0.62  
>>/home/morama/gas/pili/align/gi-21909638.pep (183 aa)  
initn: 239 init1: 164 opt: 261 Z-score: 80.5 expect(): 0.62  
Smith-Waterman score: 302; 31.5% identity in 200 aa overlap  
(14-213:2-181)

```

gi-13621430.      10      20      30      40      50      60
                  MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
gi-21909638.      :| : : :| : : :| : : :| : : :| : : :| : : :
                  MILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTP-----FSIAL

```

FIGURE 59D

PCT/US05/27239

20 30 40

70 80 90 100 110 120

gi-13621430. EALDKESPLPNSVTTTSVKGNKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVEV  
 |::| : : : : : |::|::| |::| |::| |::| : : : : : |::|::|  
 gi-21909638. ESIDAMKTIEE---ITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVL  
 50 60 70 80 90

130 140 150 160 170 180

gi-13621430. IYVLYNEQSGALETNLVSNNKLGETEKSELIFKQYSEKTPPEHPQDTEKEKPKKRNIGI  
 :|||::| : : : : : |::|::| : : : : : |::| : : : : : |::|::|  
 gi-21909638. VYVYDDED-GTLVAKVISRRAGDEEKSAITFKPKWLKPIPPROPNI-----PKTP-----  
 100 110 120 130 140

190 200 210

gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKLLKTSK  
 || :||: | : : |::|::| : : : : : |::|::|  
 gi-21909638. LPLAGEVKSLGILSIVLLGLLVLLYV-KKLSRL  
 150 160 170 180

gi-13621430.pep

/home/morana/gas/pili/align/gi-19745303.pep

gi|19745303|ref|NP\_606439.1| hypothetical protein [Streptococcus pyogenes MGAS8232]

SCORES Init1: 121 Initn: 121 Opt: 126 z-score: 61.4 E(): 6.7  
 >>/home/morana/gas/pili/align/gi-19745303.pep (344 aa)  
 initn: 121 init1: 121 opt: 126 Z-score: 61.4 expect(): 6.7  
 Smith-Waterman score: 126; 27.0% identity in 100 aa overlap  
 (59-155:84-183)

30 40 50 60 70 80

gi-13621430. LTASINIEVINQVDVATNKQSSDIDETFMFVIEALDKESPLP--NSVTTTSVKNGN-KTSF  
 |::|::| : : : : : |::| : : : : : |::|::|  
 gi-19745303. LMPKADYTFKVEADDNAKGKTKDGLDIKPGVIDGLENTKTIHYGNSDKTTAKEKSVNFDF  
 60 70 80 90 100 110

90 100 110 120 130 140

gi-13621430. EQLTFSEVGQYHYKIHQLLGKNSQYHYDETVEVVIYVLYNEQSGALETNLVSNNKLGETE  
 : : | ||| : : : : : |::| : : : : : |::| : : : : : |::|::|  
 gi-19745303. ANVKFPGVGVIYRTVSEVNGNKAGIAYDSQQWTVDVYVNNREDGGFEAKYIVSTEGGQSD  
 120 130 140 150 160 170

150 160 170 180 190 200

gi-13621430. KSELIFKQYSEKTPPEHPQDTEKEKPKKRNIGILPSTGEMVSYVSALGIVLVATITLY  
 | : : : | : : : : :  
 gi-19745303. KKPVLFRNFDDTSLKVTKKVTGNTGEHQRSFSFTLLTPNECFEKGQVNNILQGGETKK  
 180 190 200 210 220 230

gi-13621430.pep

/home/morana/gas/pili/align/gi-13621428.pep

gi|13621428|gb|AAK33238.1| hypothetical protein [Streptococcus pyogenes]

SCORES Init1: 58 Initn: 86 Opt: 122 z-score: 60.9 E(): 7.2  
 >>/home/morana/gas/pili/align/gi-13621428.pep (340 aa)  
 initn: 86 init1: 58 opt: 122 Z-score: 60.9 expect(): 7.2  
 Smith-Waterman score: 135; 29.1% identity in 172 aa overlap

FIGURE 59E

(8-159) PCT/US2005/027239

```

      10      20      30      40      50
gi-13621430. MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVD-VATNKQSSDIDETFMF-
      :| | | | | :|::|::| : : :| : :| | :| :| | | :
gi-13621428. MKLRHLLLTGAALTSFAAT-TVHGETVVNGAKLTVTKNLDLVNSNALIPNTDFTFKIE
      10      20      30      40      50

      60      70      80      90      100
gi-13621430. ---VIEALDK-----ESPLPN-SVTTSVKNGKTSFEQLTFSEV-----GQYHYKI-H
      | | :| :|::|::| :| | |::|: : :| | | | | | | | :| | :| :
gi-13621428. PDTTVNEDGNKFKGVALNTPMTKVITYNSDKGGSNTKTAEFDFSEVTFEKPQGVYVYKVT
      60      70      80      90      100      110

      110      120      130      140      150      160
gi-13621430. QLLGKNSQYHYDETVYEVVYVLYNE-QSCALETNLVSNKLGETEKSELIFKQEYSEKTP
      : : | | | | | :|::|::| : : | :| :| | | | :| | :| :| :
gi-13621428. EKIDKVPQVSYDTSYTVQVHVLWNEEQQKPVATYIVGYKEGS--KVPIQFKNSLDSTTL
      120      130      140      150      160      170

      170      180      190      200      210
gi-13621430. EPHQPDTEKEKPKQKRNGLPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
gi-13621428. TVKKKVSQGGDRSKDFNGLTLKANQYKASEKVMIEKTTKGGQAPVQTEASIDQLYHF
      180      190      200      210      220      230

```

gi-13621430.pep

/home/morama/gas/pili/align/gi-19224137.pep

gi|19224137|gb|AAL86408.1|AF447492\_5 EftLSL.A [Streptococcus pyogenes]

```

SCORES   Init1: 88   Initn: 88   Opt: 119   z-score: 60.4   E(): 7.5
>>/home/morama/gas/pili/align/gi-19224137.pep           (342 aa)
initn:   88 init1: 88 opt: 119 Z-score: 60.4 expect(): 7.5
Smith-Waterman score: 119;   29.7% identity in 111 aa overlap
(72-176:93-201)

```

```

      50      60      70      80      90
gi-13621430. DVATNKQSSDIDETFMFVIEALDKESPLPNVTSVKNGK-----TSFEQLTFSEVGQY
      | : :| :|::| :| :| | | | |
gi-19224137. SVNPDAAATGTESNLPIKPGIAVNNQDIKVSYSNTDKTSGKEKQVVVDPMKVTFFPSVGIY
      70      80      90      100      110      120

      100      110      120      130      140      150
gi-13621430. HYKIHOLLGKNSQYHYDETVYEVVYVLYNEQSCALETNLVSNKLGETEKSELI-FKQEY
      | : :| :|::| :| :| | | | |
gi-19224137. RYVVTENKGTAEQVYDDTKVLVDVYVGNNEKGG-LEPKYLVSKKGD SATKEPIQFNNSF
      130      140      150      160      170      180

      160      170      180      190      200      210
gi-13621430. SEKTPEPHQPDTEKEKPKQKRNGLPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
      | | : :| :| :| |
gi-19224137. -ETTSLKIEKEVTGNTGDHKAFTFTLTLPNEYEASSVVKIEENGQTKDVKIGEAYKF
      190      200      210      220      230      240

```

gi-13621430.pep

/home/morama/gas/pili/align/gi-50913503.pep

gi|50913503|ref|YP\_059475.1| Fibronectin-binding protein [Streptococcus pyogenes MGAS10394]

FIGURE 59F

PCT/US05/27239

SCORES Init1: 73 Initn: 73 Opt: 117 z-score: 60.4 E(): 7.6  
>>/home/morana/gas/pili/align/gi-50913503.pep (627 aa)  
initn: 73 init1: 73 opt: 117 Z-score: 60.4 expect(): 7.6  
Smith-Waterman score: 118; 28.7% identity in 87 aa overlap  
(129-215:549-625)

```
100 110 120 130 140 150
gi-13621430. KIHQLLGKNSQYHYDETVEVVIYVLYNEQSGALETNLVSNKLGETEKSSELIFFKQYSEK
gi-50913503. IETEDTKEPEVLMGGQSESVEFTKDTQTGMSGFSETATVV---EDTRPKLVFHFDDNNEP
120 130 140 150 160 170
gi-13621430. TPEPHQPDTEKEKPKQKRNGLPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSEK
gi-50913503. KVEEN---REKPTKNITPILPATGDIENVLAFLGILILSVLSIFSLKKNKQSNKKV
180 190 200 210
gi-13621430. TPEPHQPDTEKEKPKQKRNGLPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSEK
gi-50913503. KVEEN---REKPTKNITPILPATGDIENVLAFLGILILSVLSIFSLKKNKQSNKKV
```

gi-13621430.pep

/home/morana/gas/pili/align/gi-19224134.pep

gi|19224134|gb|AAL86405.1|AF447492\_2 protein F [Streptococcus pyogenes]

SCORES Init1: 73 Initn: 73 Opt: 115 z-score: 60.1 E(): 7.8  
>>/home/morana/gas/pili/align/gi-19224134.pep (698 aa)  
initn: 73 init1: 73 opt: 115 Z-score: 60.1 expect(): 7.8  
Smith-Waterman score: 115; 27.4% identity in 73 aa overlap  
(143-215:631-697)

```
120 130 140 150 160 170
gi-13621430. DETVEVVIYVLYNEQSGALETNLVSNKLGETEKSSELIFFKQYSEKTPPEPHQPDTEKEK
gi-19224134. VLMGGQSESVEFTKDTQTGMSGFSETVTIVEDTRPKLVFHFDDNNEPKVEEN---REK
180 190 200 210
gi-13621430. POKKRNGLPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSEK
gi-19224134. PTKNITPILPATGDIENVLAFLGILILSVLSIFSLKKNKQSNKKV
```

! Distributed over 1 thread

! Start time: Wed Sep 15 18:45:36 2004

! Completion time: Wed Sep 15 18:45:38 2004

! CPU time used:

! Database scan: 0:00:00.1

! Post-scan processing: 0:00:00.3

! Total CPU time: 0:00:00.4

! Output File: gi-13621430.fasta

FIGURE 59G

**TYPE 3 pilus motifs**  
**protein F2 like fibronectin-binding protein**  
 Length: 696-733  
 LPXTG  
 pilin motif consensus PK (X<sub>7</sub>) K  
 E box consensus ETxAPxGY

SpyM3_0104/21909640	pilin motif	155 PKEKPIIYFK
	E box	398 YTFVETAAPDGY
SPs0106/28895018	pilin motif	269 PKEKPIIYFK
	E box	512 YTFVETAAPDGY
SpyM18_0132/19745307	pilin motif	269 PKEKPIIYFK
	E box	512 YTFVETAAPDGY
orf84	pilin motif	269 PKEKPIIYFK
	E box	512 YTFVETAAPDGY

**TYPE 4 pilus motifs**  
**protein F2 like fibronectin-binding protein**  
 Length: 1161  
 LPXTG  
 pilin motif consensus PK (X<sub>7-8</sub>) K  
 E box consensus YxLxETxAPxGY

The protein is longer than the proteins belonging to TYPE 3 and has 4 possible pilin motifs and 2 E boxes

19224141	pilin motifs	215 PKGISQDIPK
		571 PKGYQQVTEK
		156 PKMSVVSKYK
		674 PKYDAKNQEVK
	E boxes	563 YDLYETKAPKGY
		940 YTFVETAAPDGY

FIGURE 60



Formation of pili structures on GBS appears to be correlated to FACS values for surface expression of GBS80 protein

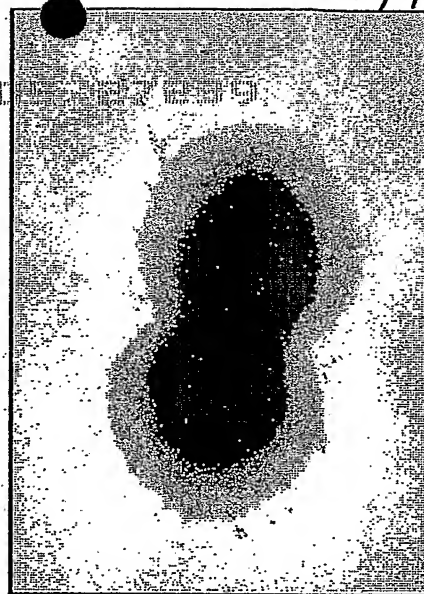
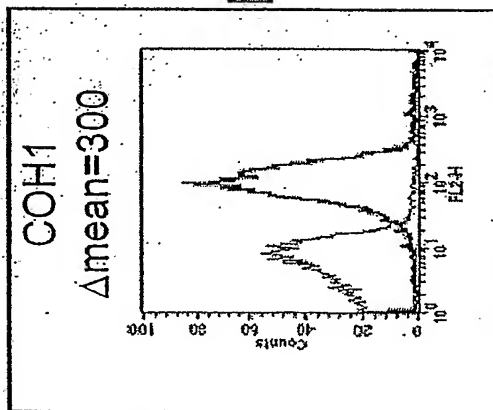
WO 2006/078318

PCT/US2005/027239

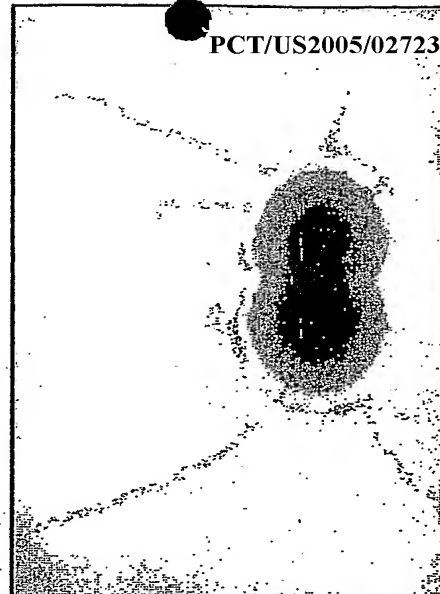
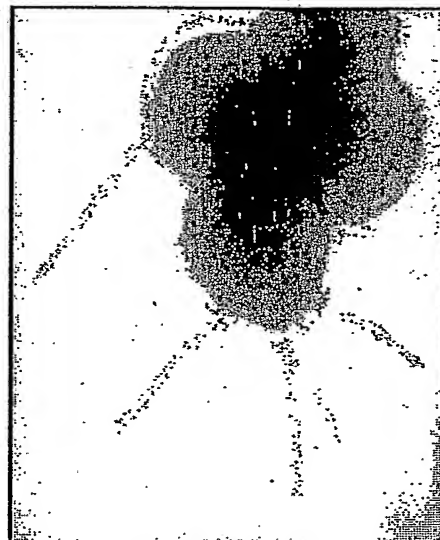
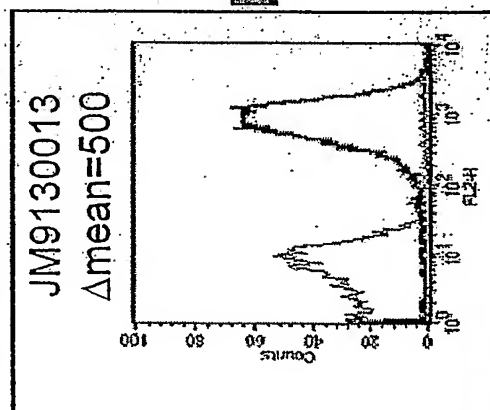
211/487

Immunogold Electron Microscopy

$\alpha$ -GBS80



$\alpha$  GBS80



PCT/US2005/027239

CHIRON VACCINES

FIGURE 61

# Surface exposure is capsule-dependent for GBS 322 but not for GBS 80

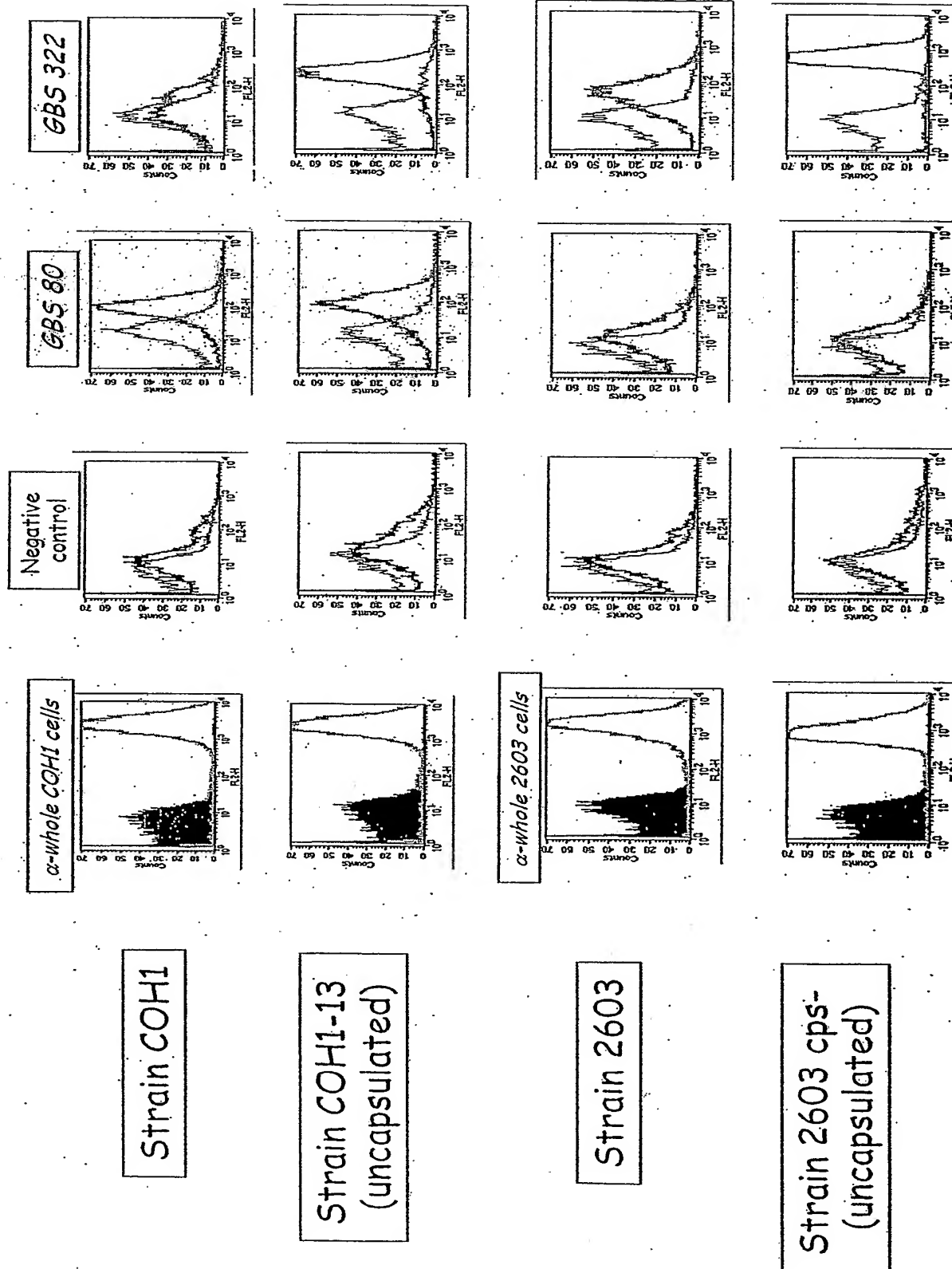
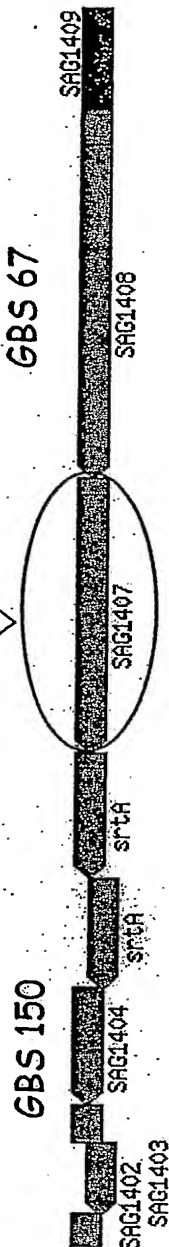


FIGURE 62

Adhesin island 2-

Operon gbs 67, 59, 150



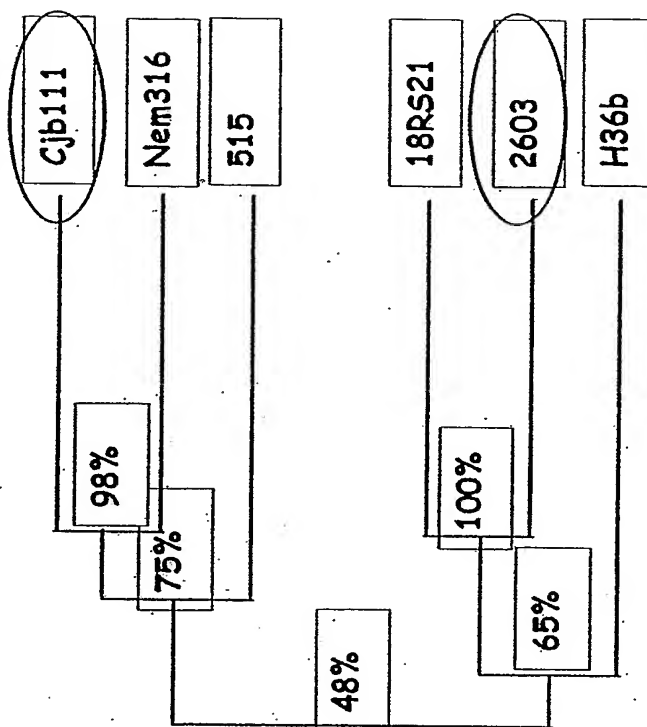
• 30 % identity with GBS 80

• By comparing GBS 59 amino acid sequence of 2603 with that of other sequenced genomes, the following homologies are obtained:

2603	
nem316	47%
h36b	62%
515	48%
cjb111	48%
18rs21	100%
coh1	not present (Spb1)
A909	not present (Spb1)

• CGH: 1/20 GBS strains analyzed (18RS21)

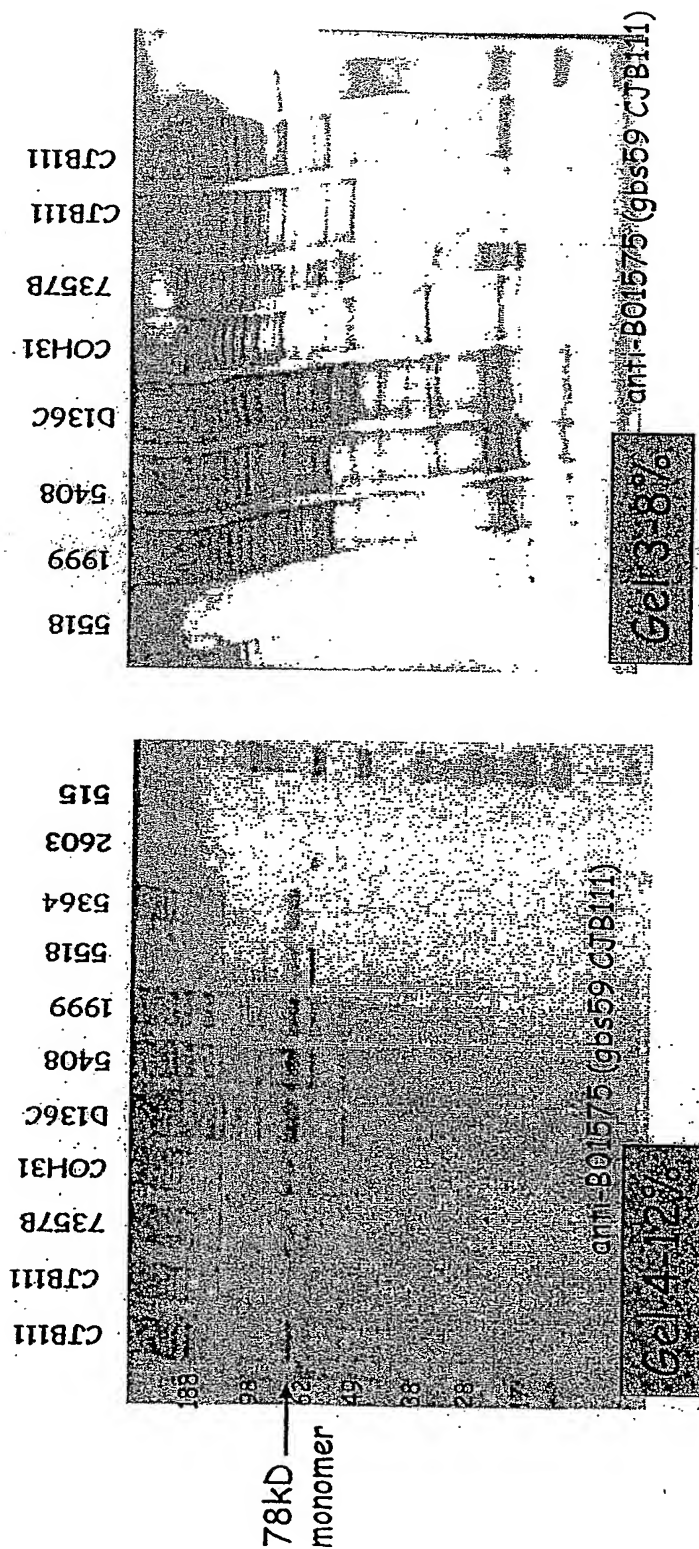
Two-by-two amino acid sequence comparison



There seem to be two clusters suggesting the presence of two major isoforms

FIGURE 63

# Western blotting with whole extracts derived from GBS strains



GBS 59 is part of a high molecular weight polymer (pilus)  
in GBS strains: CJB111, 73578, coh31, d1363c, 5408, 1999,  
5364, 5518, 515

FIGURE 64

# Western blotting with purified proteins and whole extracts derived from GBS strains

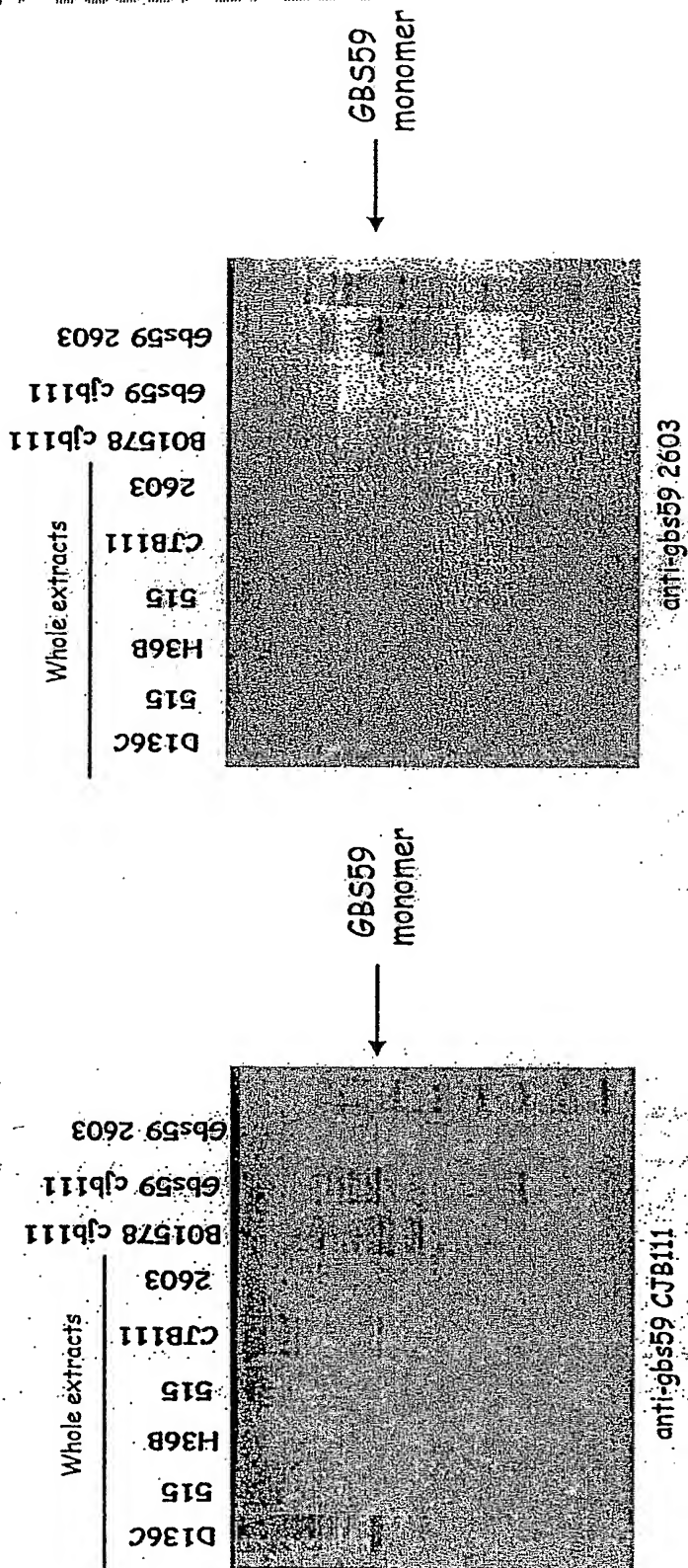
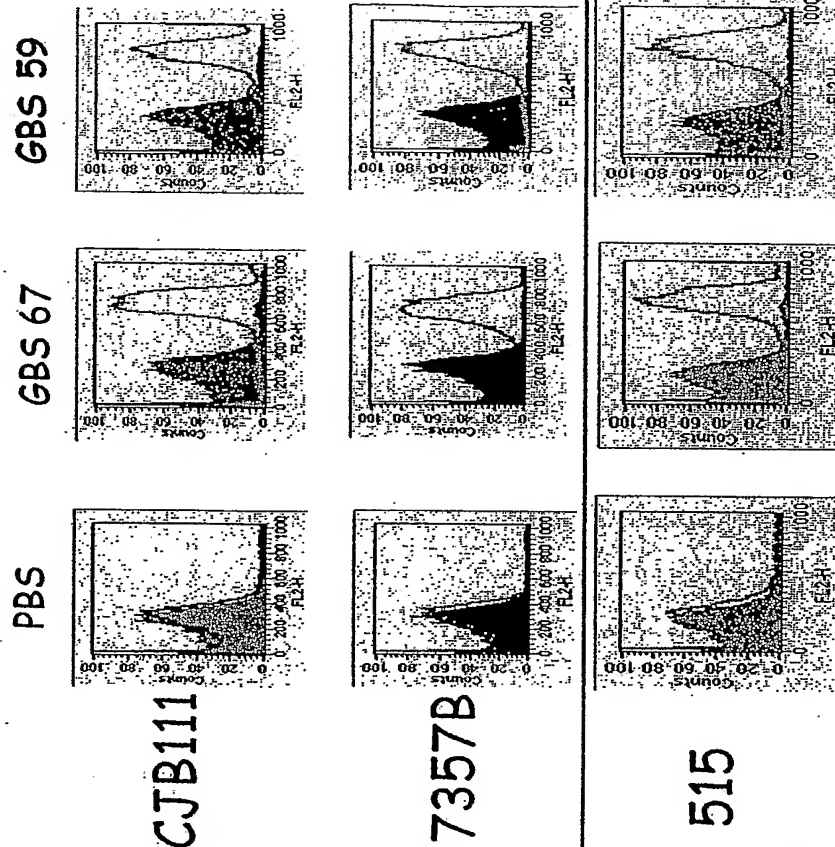


FIGURE 65

# FACS analysis using mouse antiserum after immunization with BO1575 (gbs59) from CJB111 genome

GBS strains	Type	GBS 59
DK1	Ia	565
DK8		559
Davis		577
515		583
2986		443
5551		524
7357b-		596
5518		190
D136C	III	504
COH31		505
DK21	II	249
CJB111	V	493
5364		593
2110		590
1999		594
2210		636
5408		537
1169		227



Where present GBS 59 is a highly exposed on the GBS surface

FIGURE 66



# Opsonophagocytosis assays: B01575 (gbs59-cjb111)

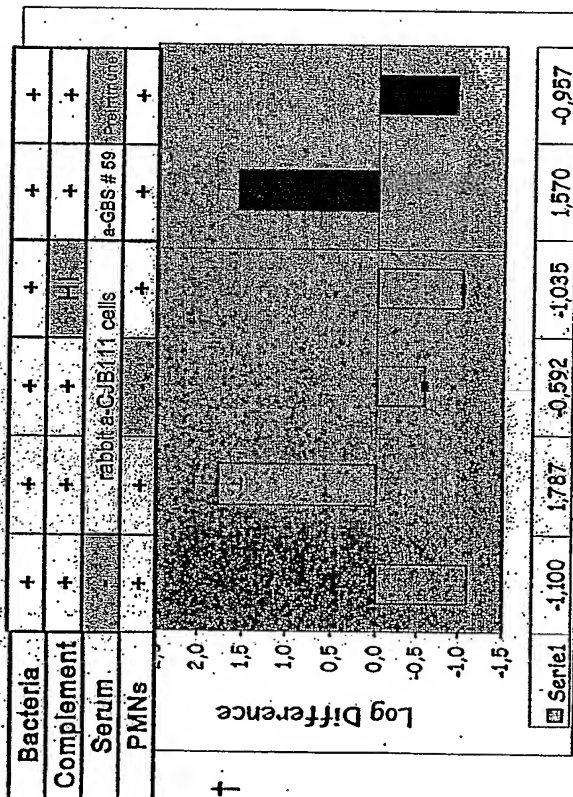
WO 2006/078318

217/487

PCT/US05/27239

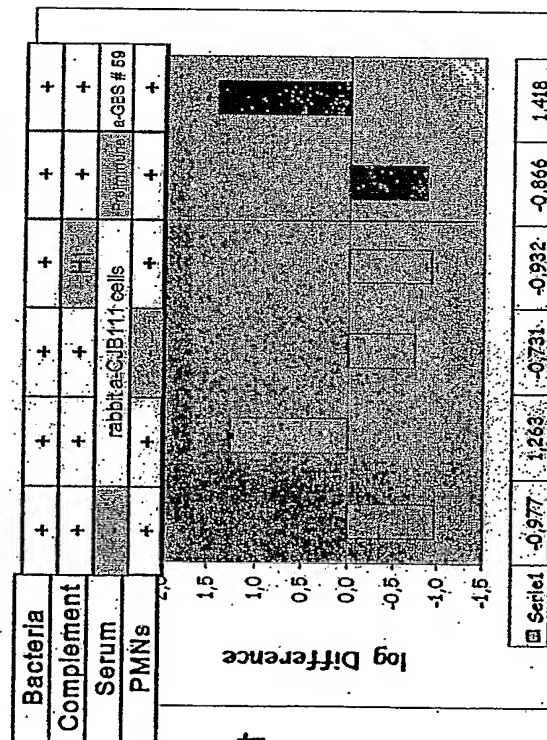
PCT/US2005/027239

I  
experiment



- CJB111 GBS strain type Ia
- Baby rabbit complement
- Human PMNs
- Positive control: anti-type V cells (rabbit serum anti fixed type V cells)

II  
experiment



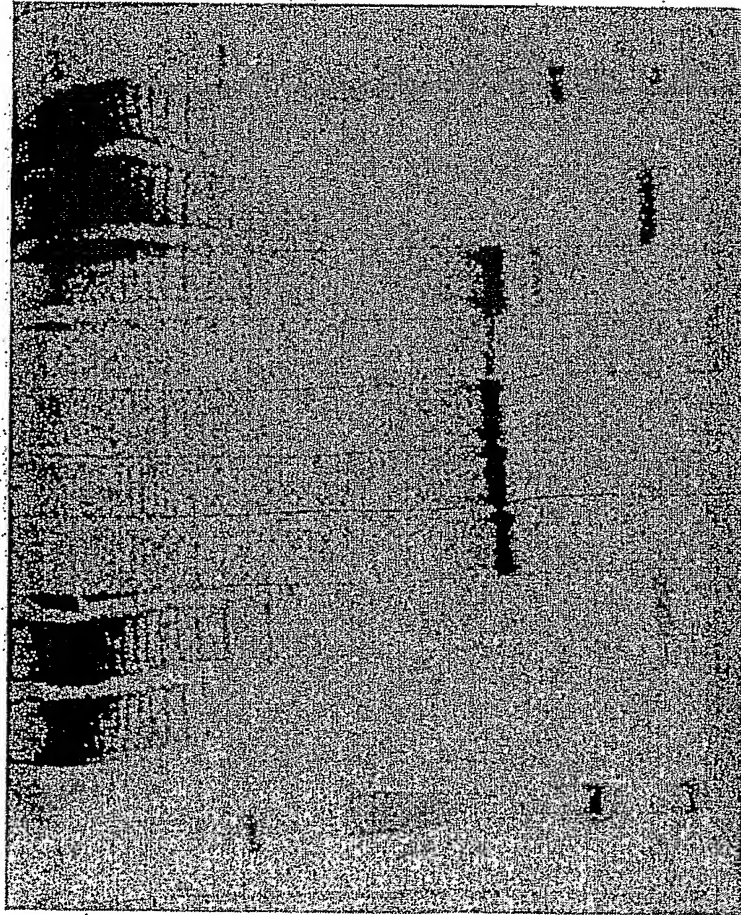
Antibodies against B01575 (cjb111) are opsonic for cjb111 GBS strain serotype V

FIGURE 67

# Association GBS 80-104

## WB $\alpha$ -80, $\alpha$ -104 JM9130013 Total Extract

Mab  $\alpha$ -80 A4/77  
 Mab  $\alpha$ -80 19G4/78  
 Mab  $\alpha$ -104 15H3/49  
 Mab  $\alpha$ -104 12A7/67  
 Mab  $\alpha$ -104 H2/32  
 Mab  $\alpha$ -104 14B3/73  
 $\alpha$ -104 POLIC.  
 $\alpha$ -80 POLIC.  
 Mab  $\alpha$ -80 19G1/77

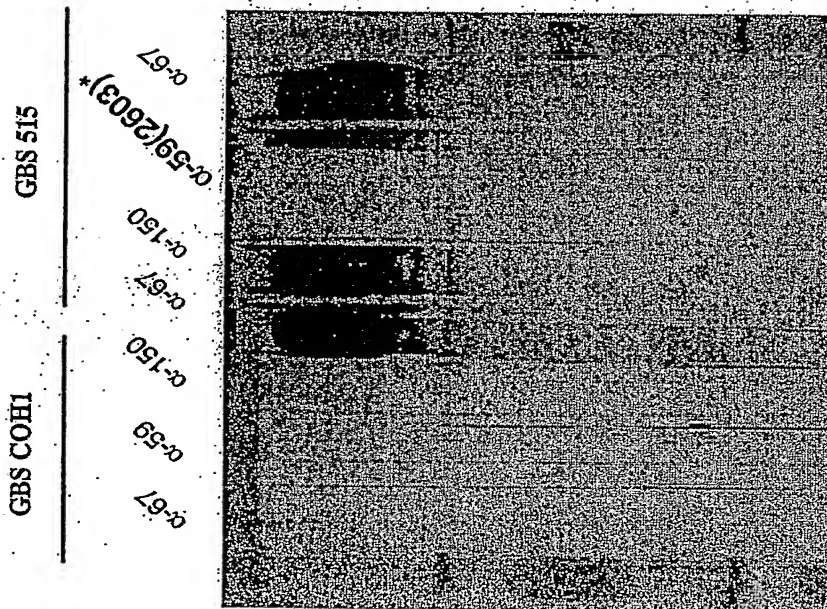


FACS	( $\Delta$ MEAN)
GBS 80	597
GBS 104	446

FIGURE 68

# WB GBS 515 Total Extract

$\alpha$ -67;  $\alpha$ -150;  $\alpha$ -59 (2603)



Controls:  
GBS COH1 total extracts

- anti-gbs59 mouse serum after immunization with SAG1407 (GBS 59) from 2603 genome

GBS 67 and GBS 150 are parts of a high molecular weight polymer (pilus) in 515 GBS strain

FIGURE 69

# Western Blotting ko GBS67 from 515 genome (clone 1.45)

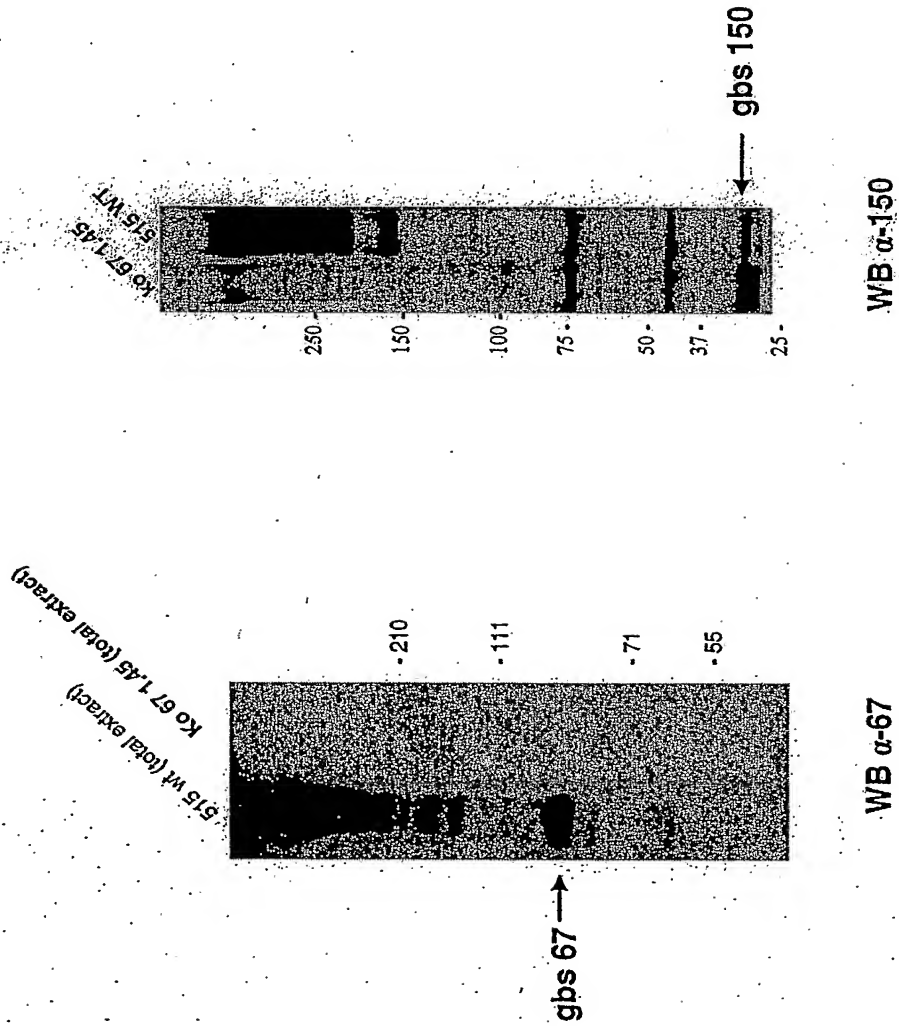


FIGURE 70

# FACS GBS 515 Δ67

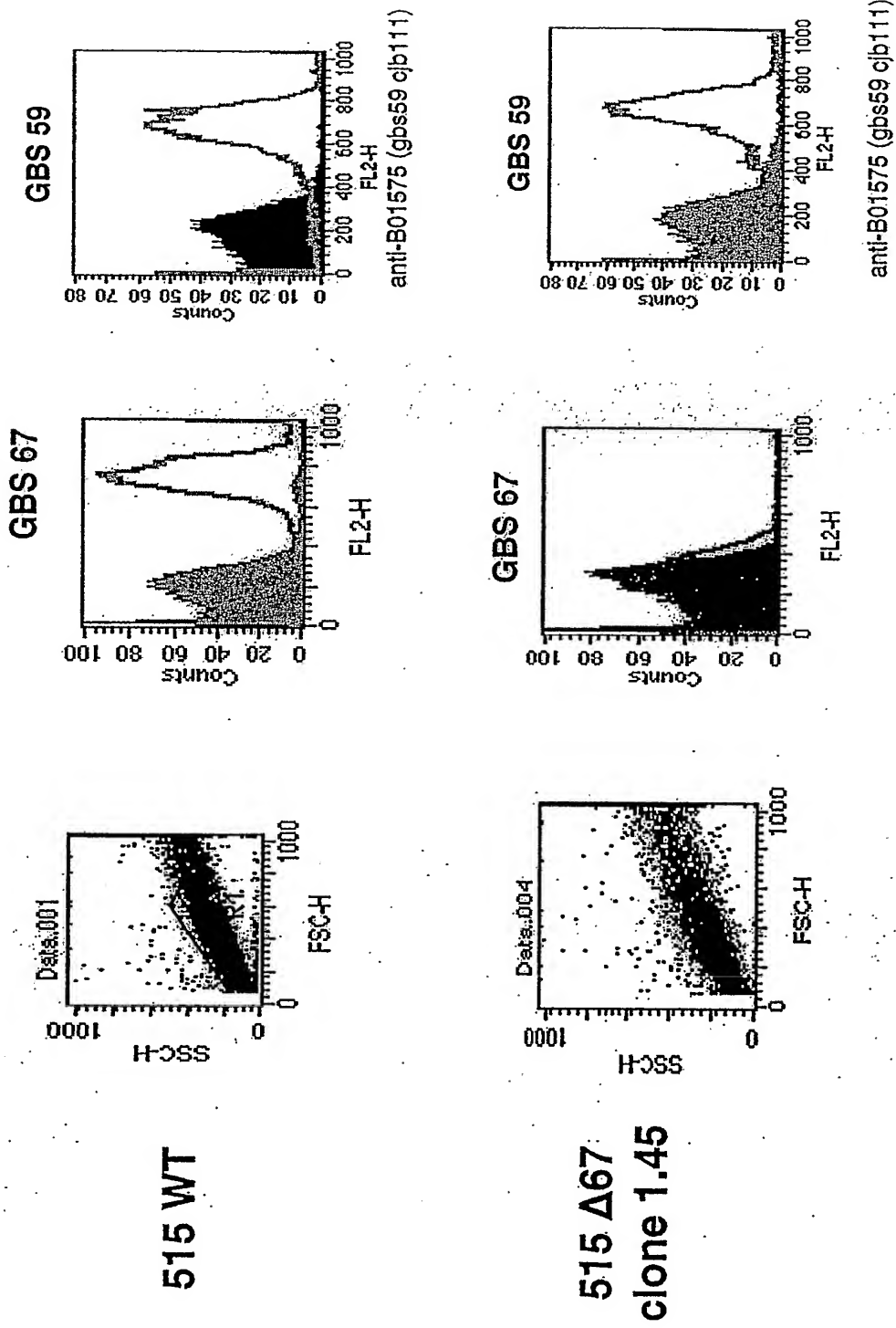


FIGURE 71

Complementation of GBS 515 KO 67 with pAM401-gbs80

GBS 80 forms a high molecular weight complex (pilus) in absence of GBS 67

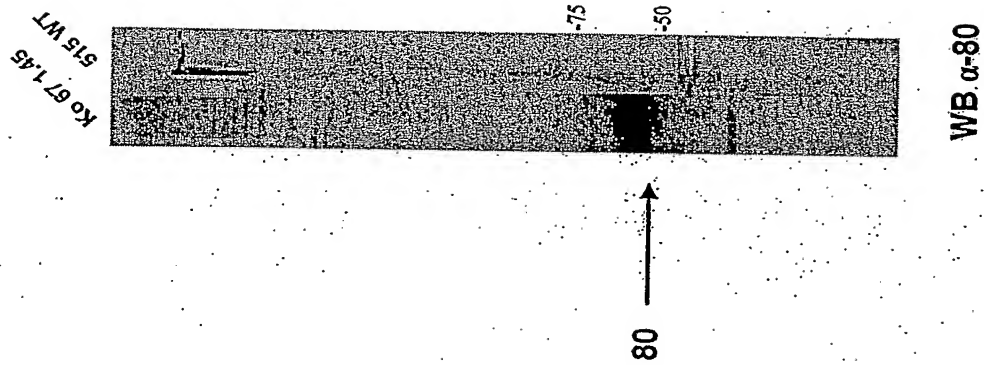


FIGURE 72



# spyM6\_0159 type 1 pilus present in M6

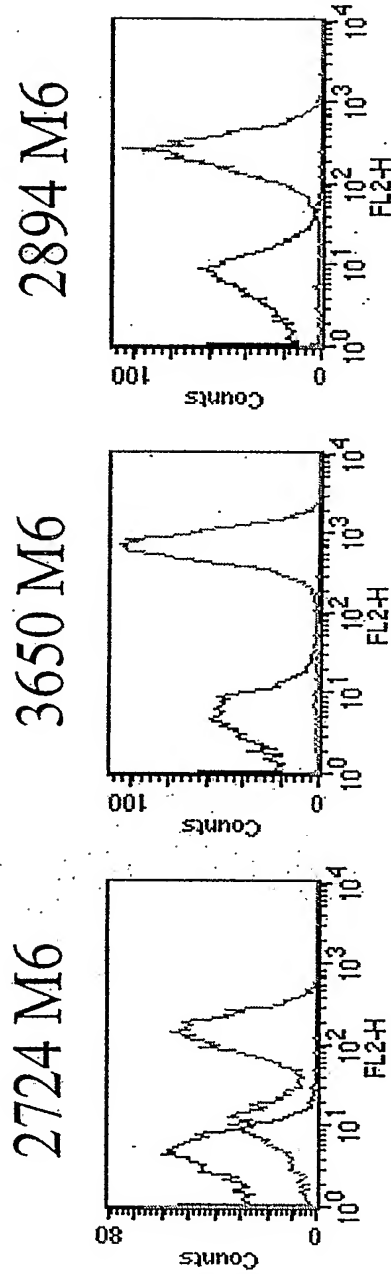


Figure 73

PCT/US05/27239 224/487

# spyM6\_0160 type 1 pilus present in M6

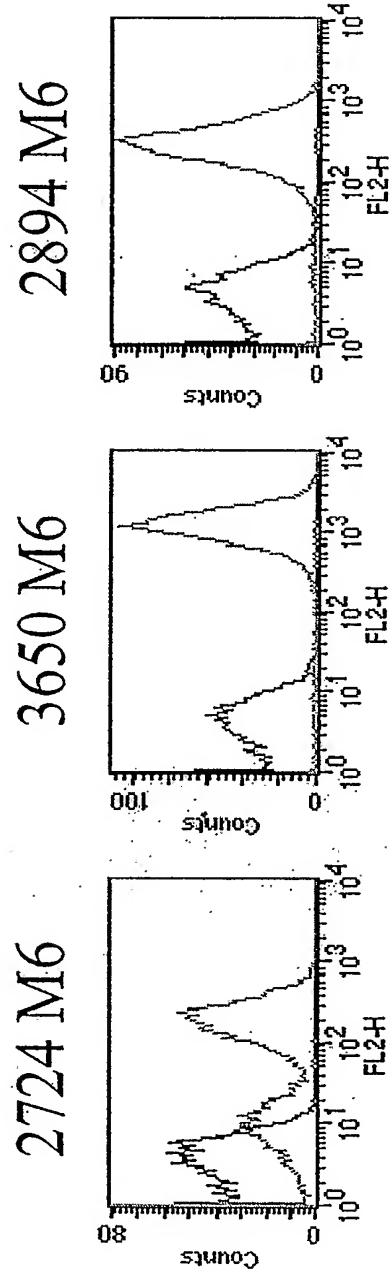


Figure 74

# Gas15 type 2 pilus present in M1

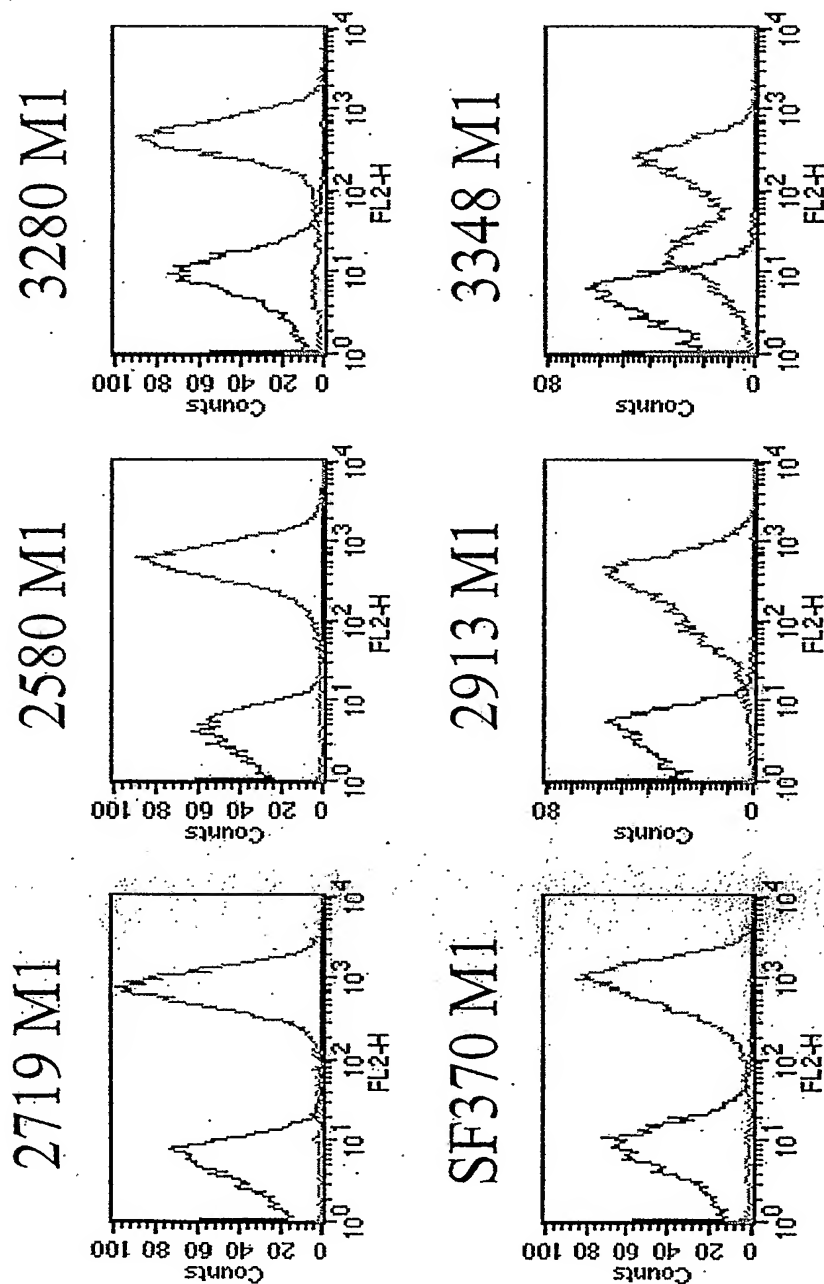


Figure 75

# Gas16 type 2 pilus present in M1

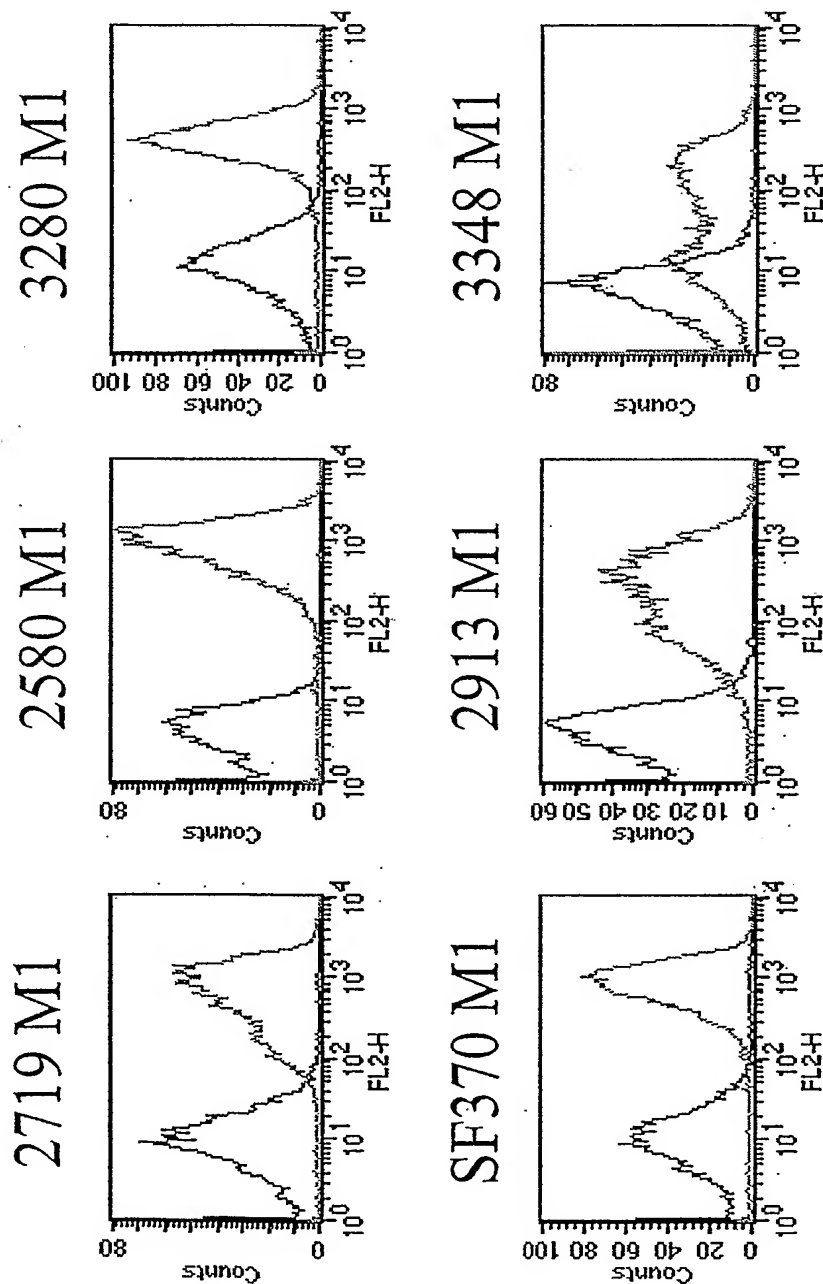


Figure 76

# Gas18 serum 1 type 2 pilus present in M1

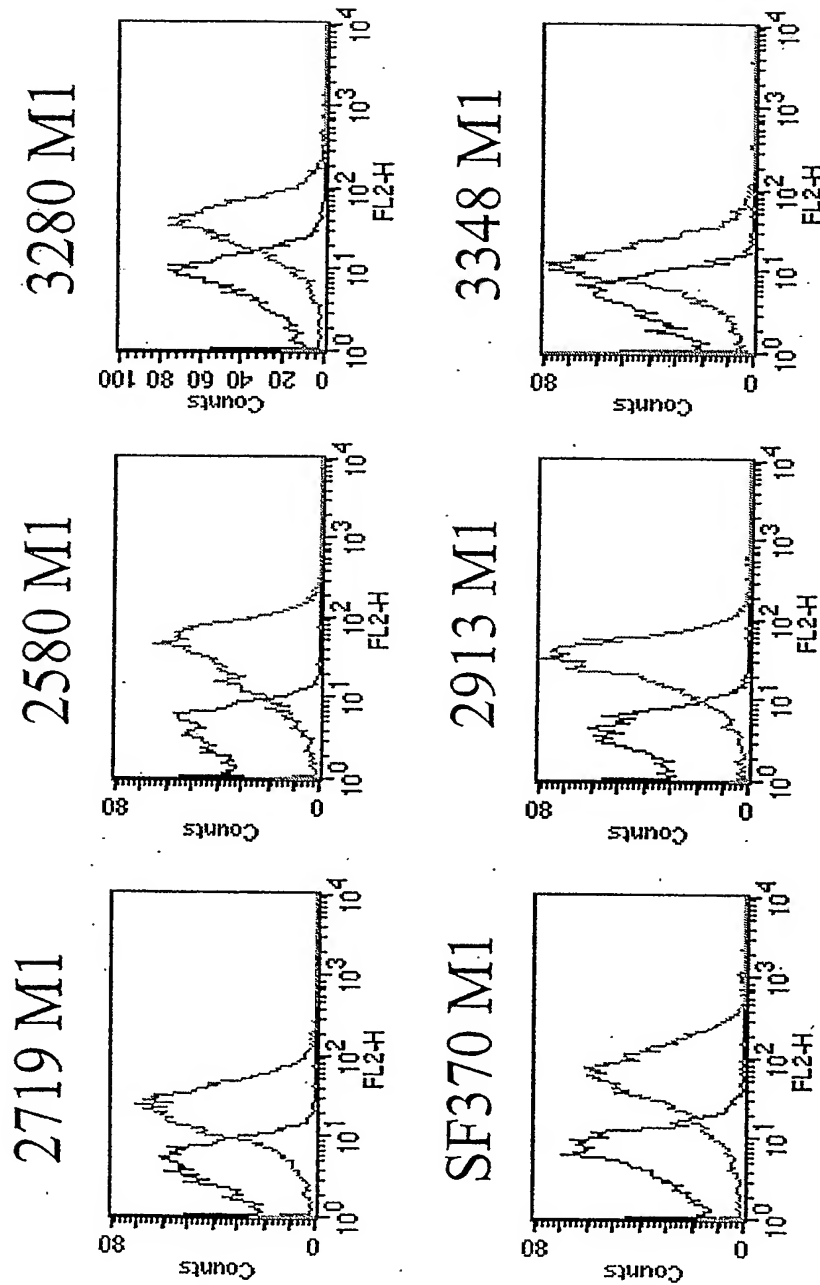


Figure 77

# Gas18 serum 2 type 2 pilus present in M1

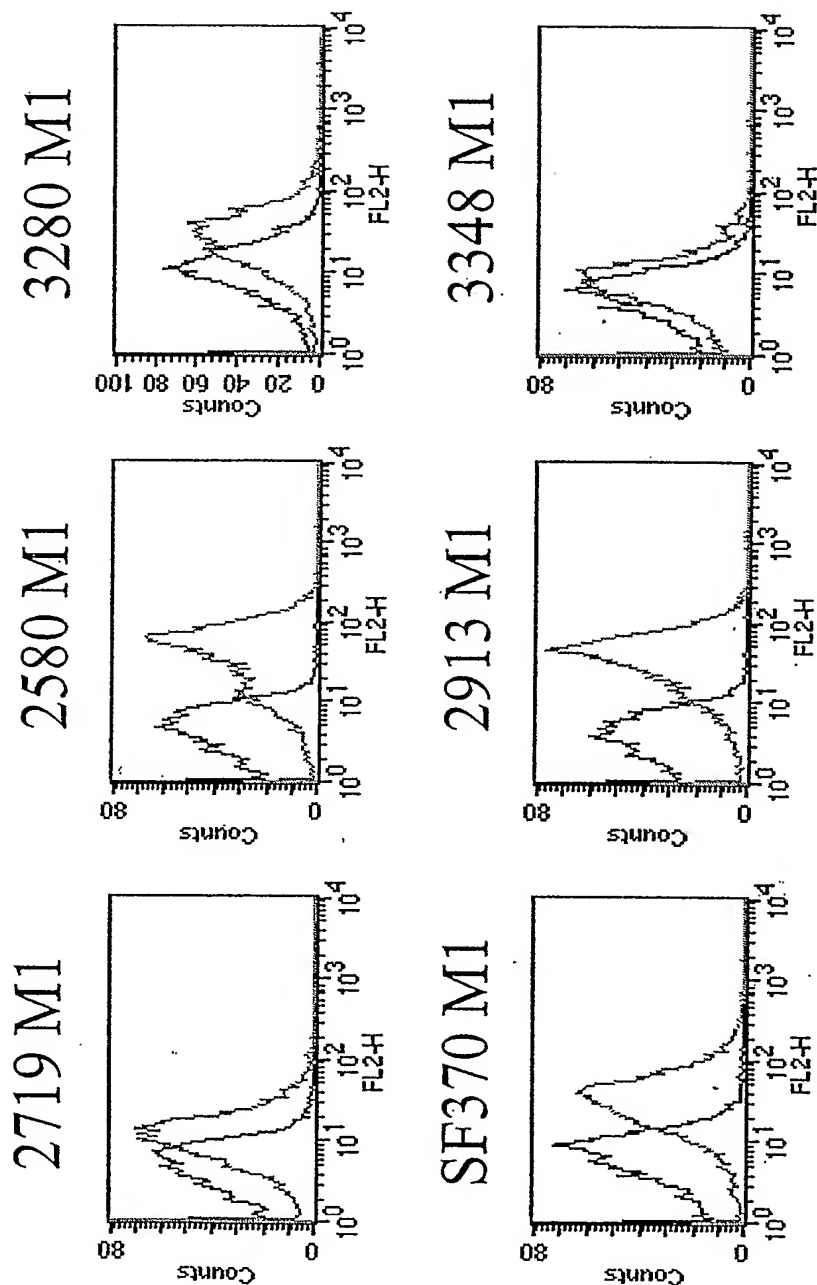


Figure 78



229/487

# Gas16p2 type 2 pilus present in M1

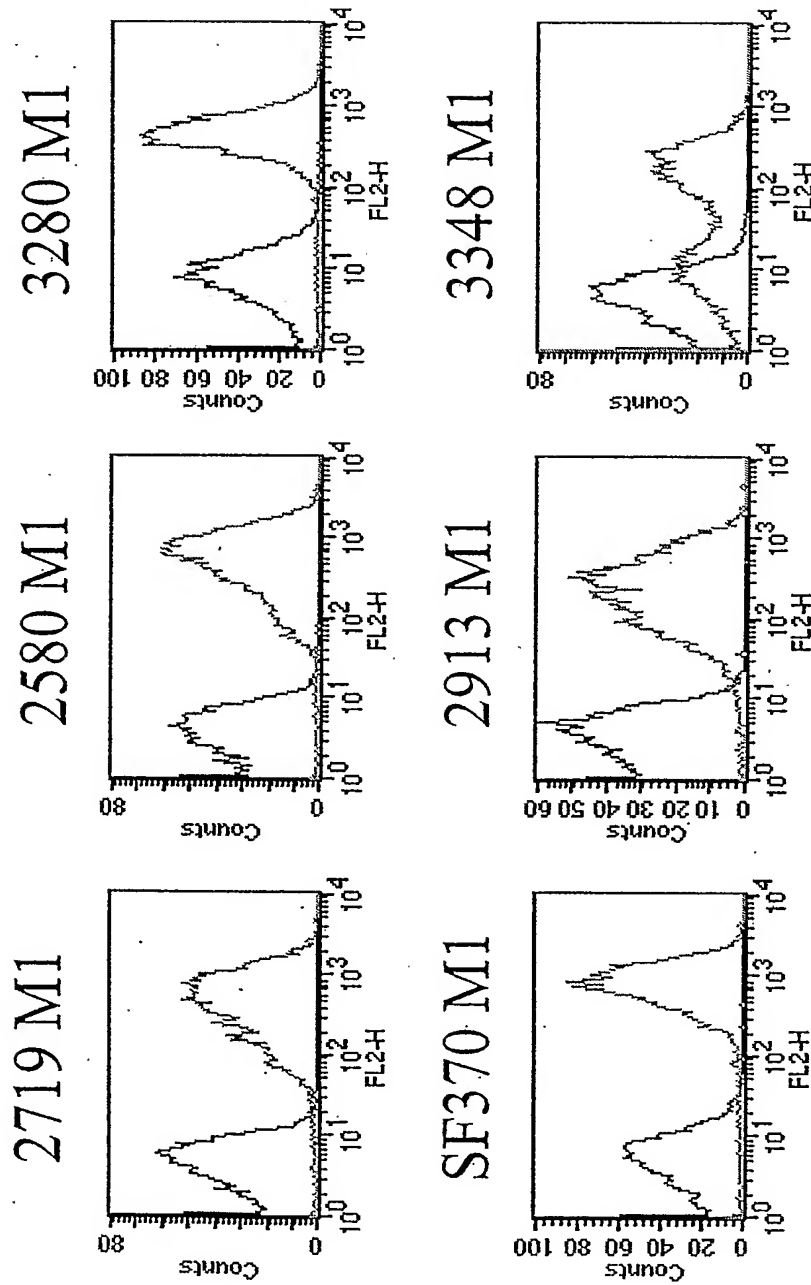


Figure 79

# spyM3\_0098 type 3 pilus present in M3

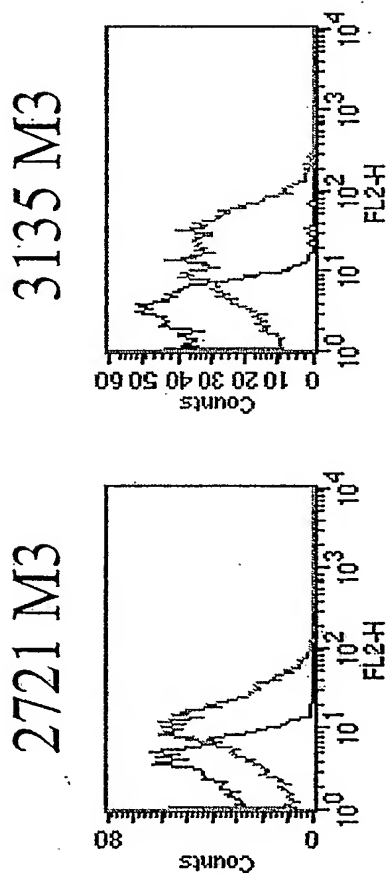
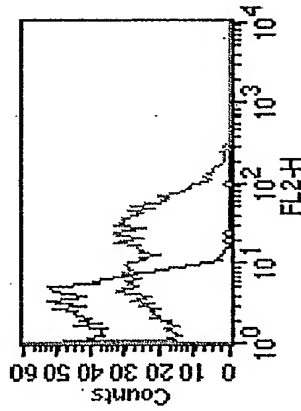


Figure 80

# spyM3\_0100 type 3 pilus present in M3

3135 M3



2721 M3

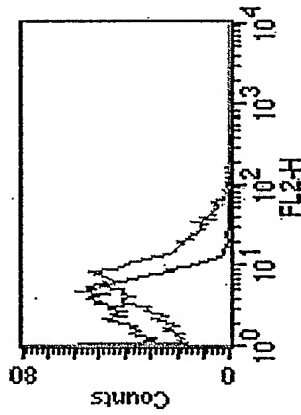
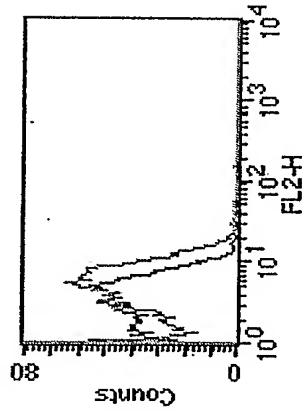


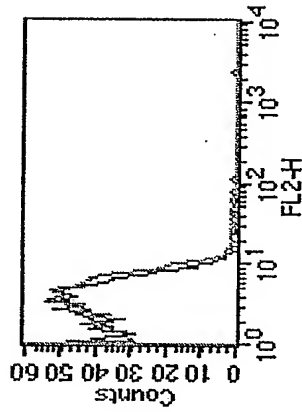
Figure 81

# spyM3\_0102 type 3 pilus present in M3

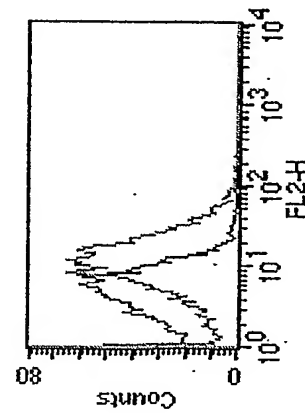
2721 M3



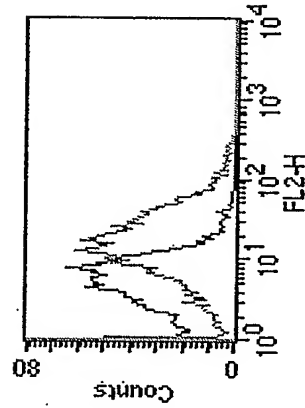
3135 M3



2724 M6



3650 M6



2894 M6

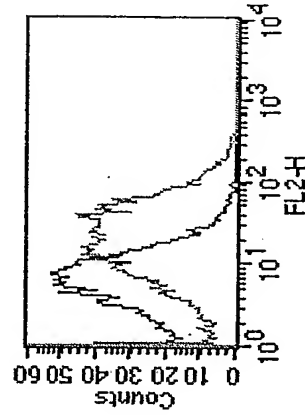


Figure 82

PCT/US05/27239 233/487

# spyM3\_0104 type 3 pilus present in M3

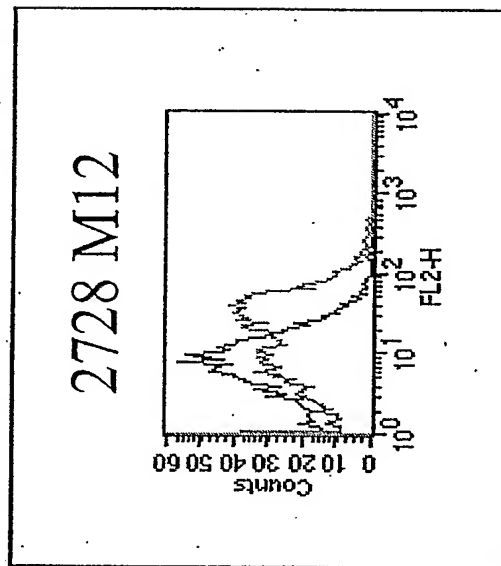
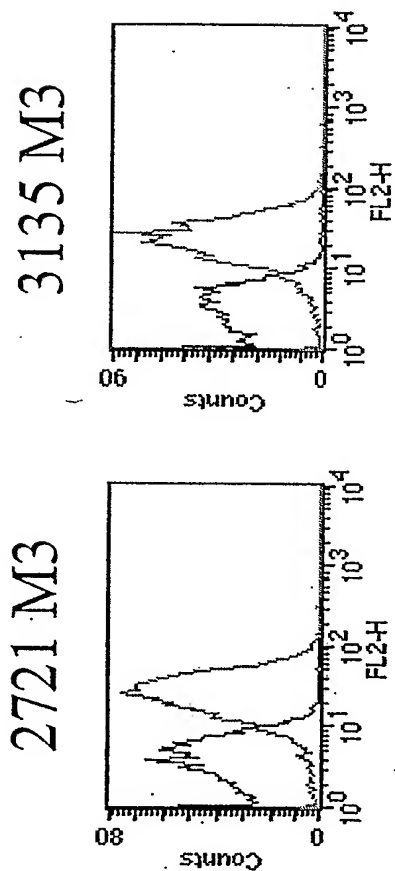
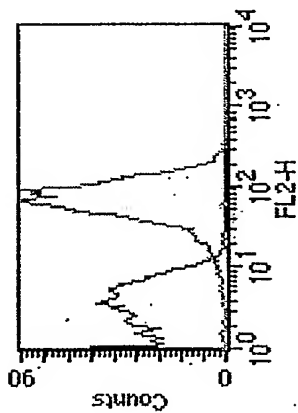


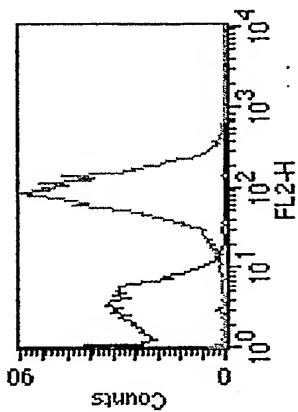
Figure 83

# spyM3\_0106 type 3 pilus present in M3

2721 M3



3135 M3



2728 M12

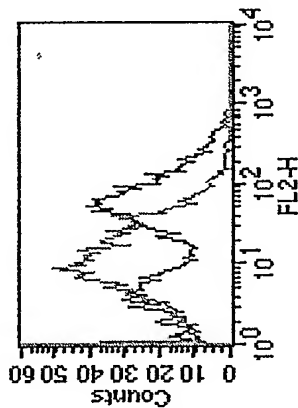


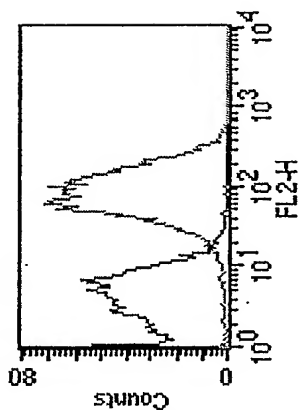
Figure 84



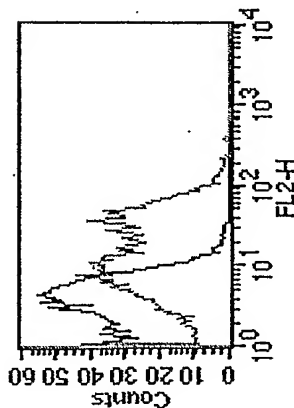
PCT/US05/27239 235/487

# 19224134 type 4 pilus present in M12

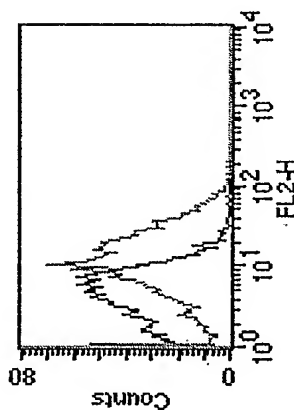
2728 M12



2724 M6



3650 M6



2894 M6

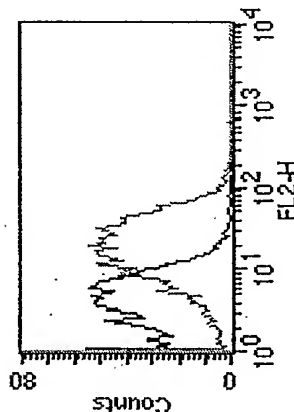


Figure 85

# 19224135 type 4 pilus present in M12

2728 M12

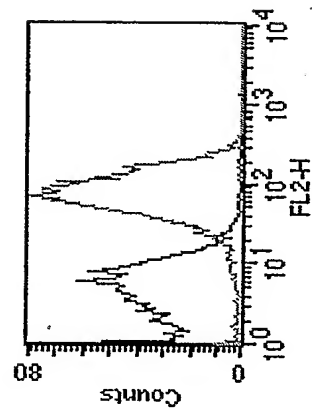


Figure 86

19224137 type 4 pilus present in M12

2728 M12

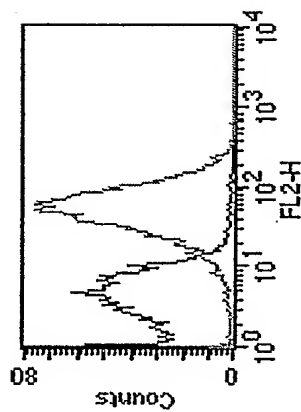


Figure 87

# 19224141 type 4 pilus present in M12

2728 M12

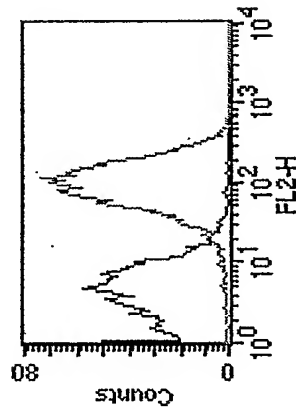


Figure 88

Figure 89

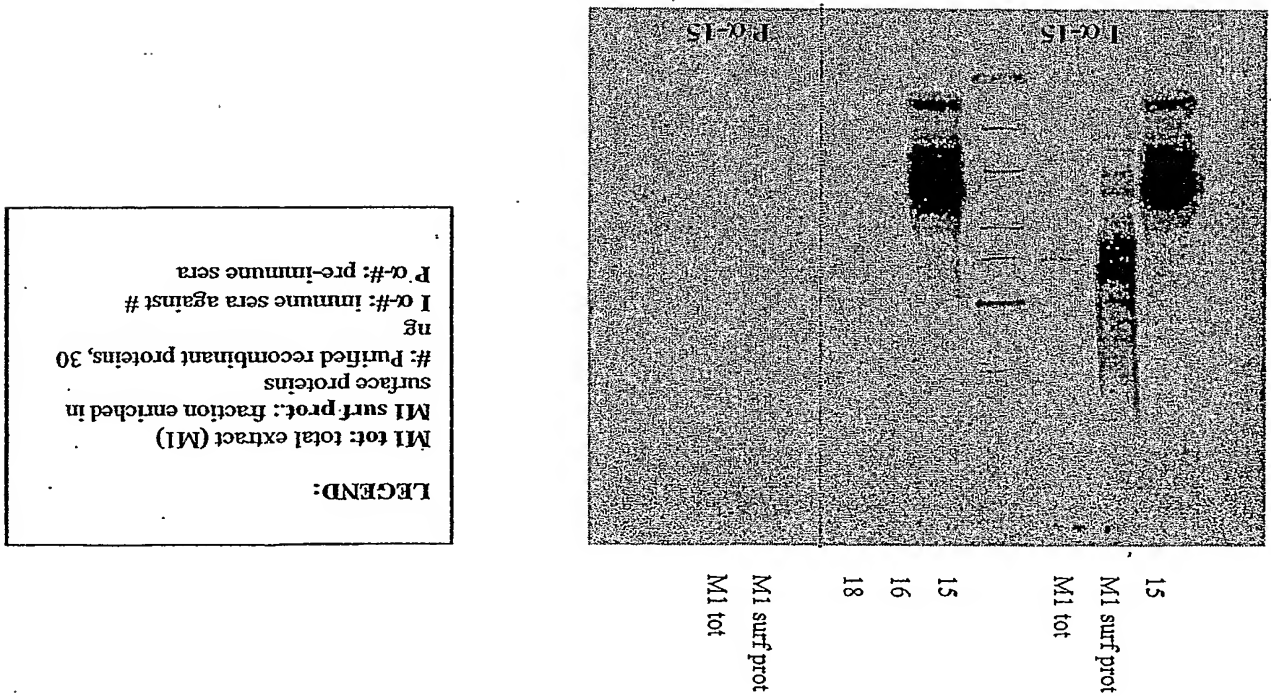
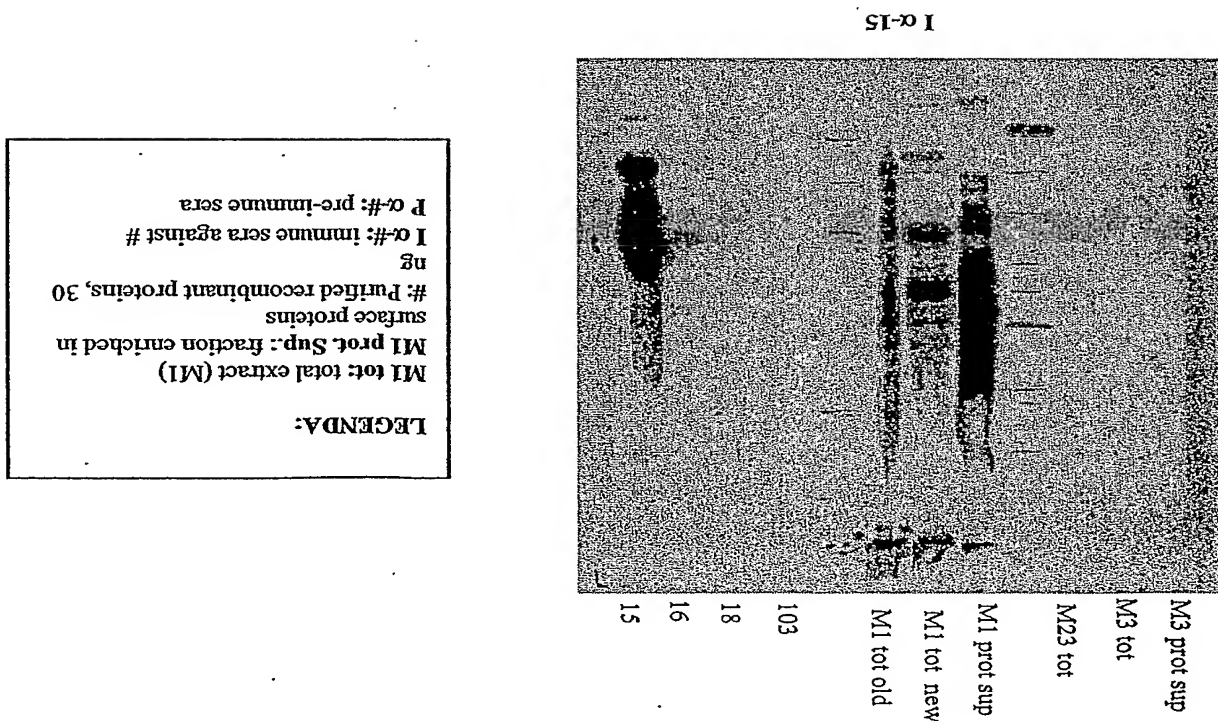


Figure 90





PCT/US05/27239 341/487

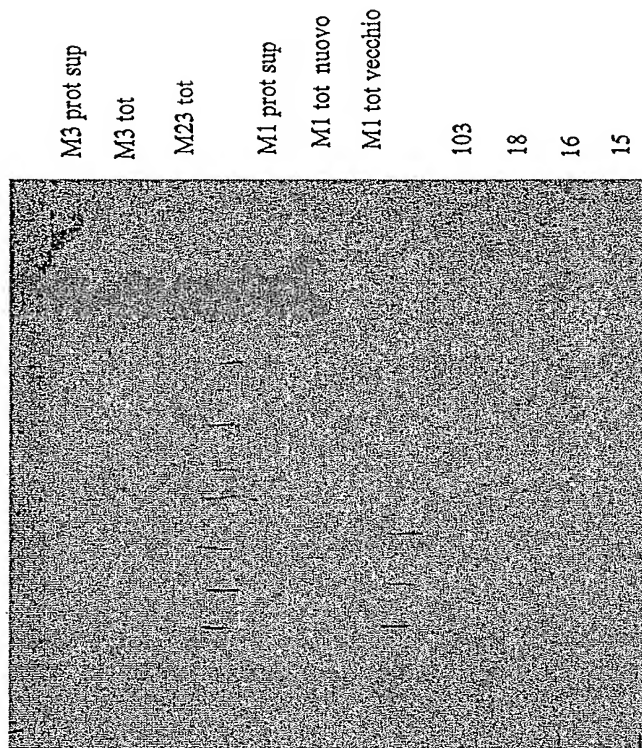
**LEGENDA:****M1 tot:** total extract (M1)**M1 prot. Sup.:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I α-#:** immune sera against #**P α-#:** pre-immune sera**P α-15**

Figure 91

PCT/US05/27239

242/487

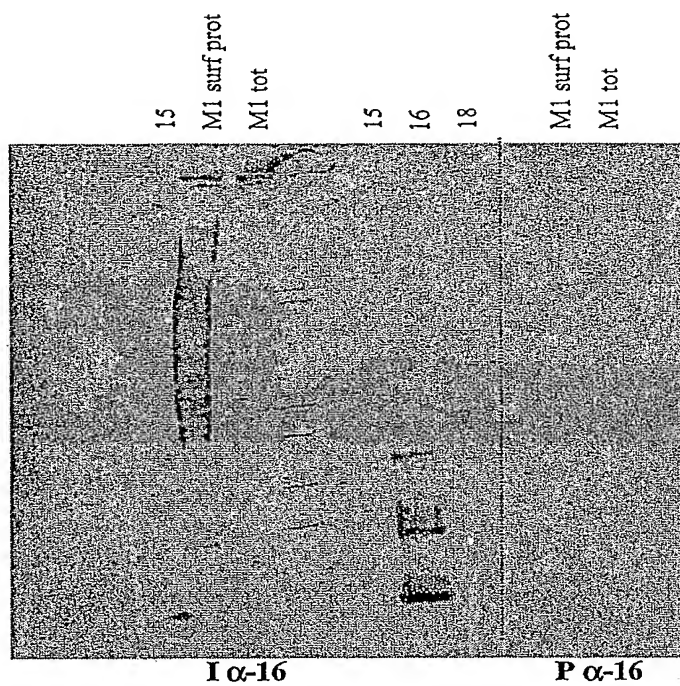
**LEGEND:****M1 tot:** total extract (M1)**M1 surf prot.:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I α-#:** immune sera against #**P α-#:** pre-immune sera

Figure 92

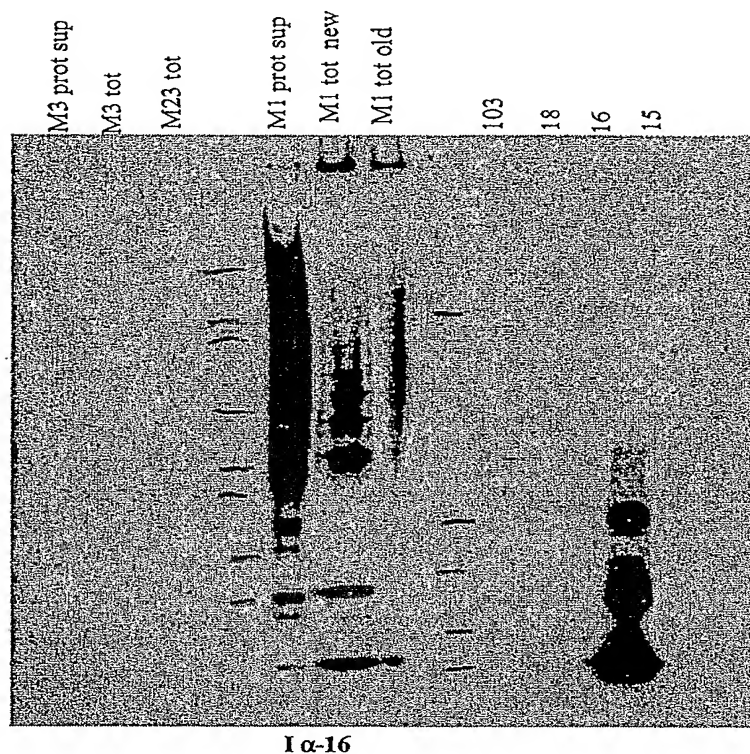
**LEGENDA:****M1 tot:** total extract (M1)**M1 prot. Sup.:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I  $\alpha$ -#:** immune sera against #**P  $\alpha$ -#:** pre-immune sera

Figure 93

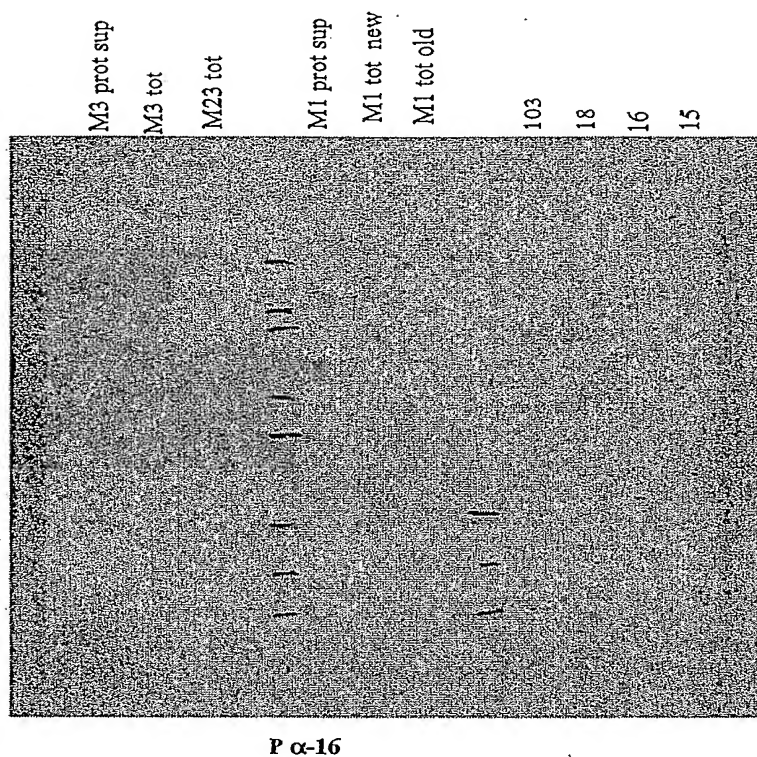
**LEGENDA:****M1 tot:** total extract (M1)**M1 prot. Sup.:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I α-#:** immune sera against #**P α-#:** pre-immune sera

Figure 94

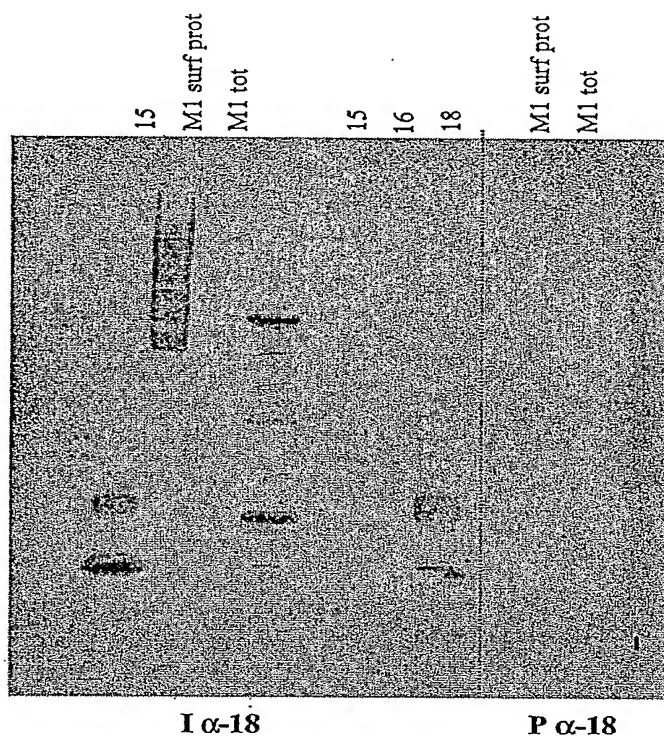
**LEGEND:****M1 tot:** total extract (M1)**M1 surf prot.:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I  $\alpha$ -#:** immune sera against #**P  $\alpha$ -#:** pre-immune sera

Figure 95

PCT/US05/27239 246/487

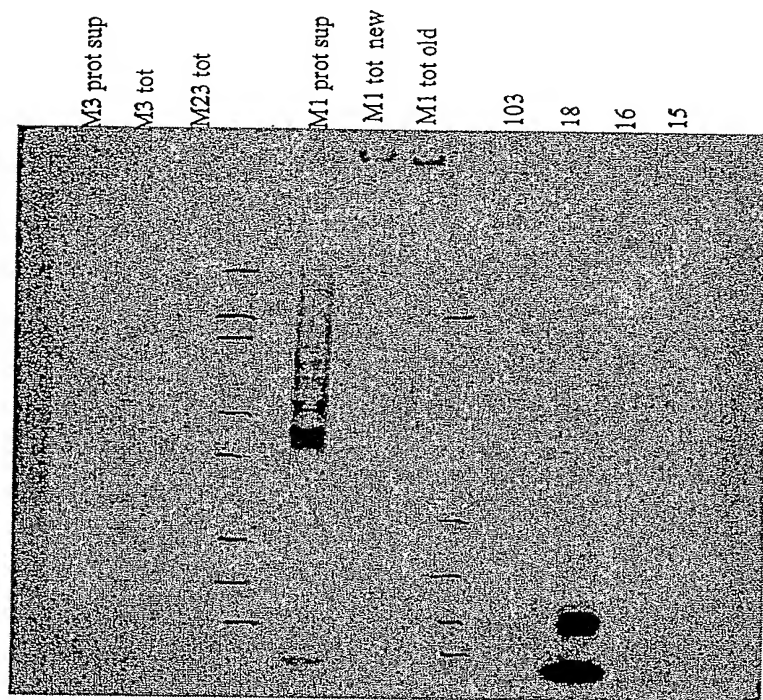
I  $\alpha$ -18**LEGENDA:****M1 tot:** total extract (M1)**M1 prot. Sup.:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I  $\alpha$ -#:** immune sera against #**P  $\alpha$ -#:** pre-immune sera

Figure 96



PCT/US2005/027239 247/487

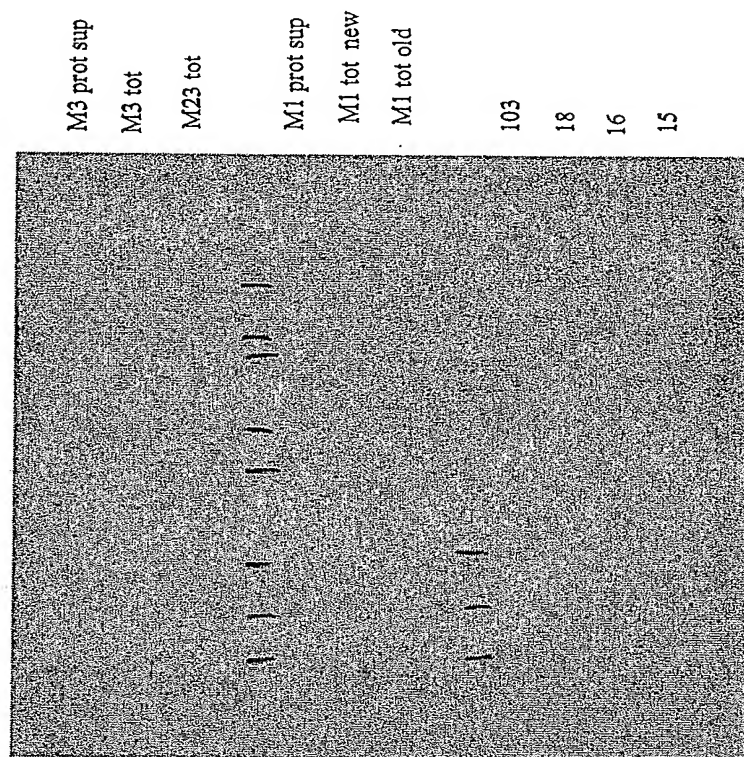
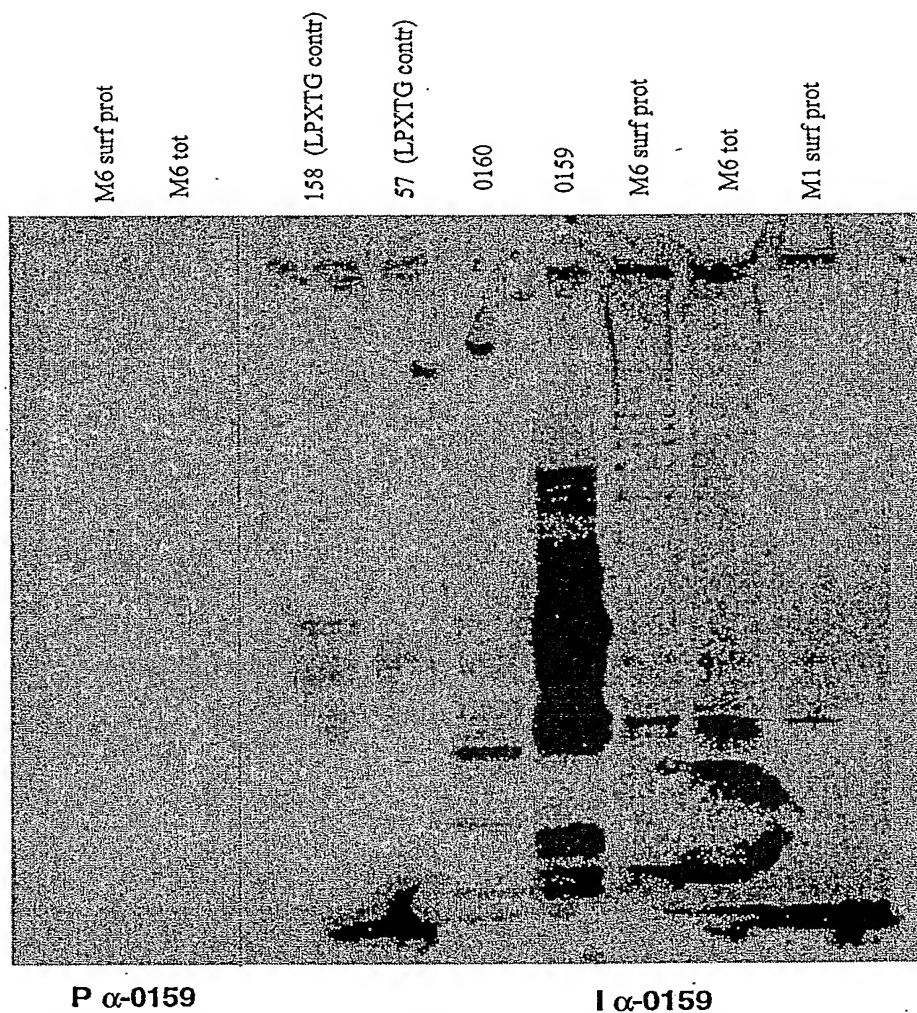
**LEGENDA:****M1 tot:** total extract (M1)**M1 prot. Sup.:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I α-#:** immune sera against #**P α-#:** pre-immune sera**P α-18**

Figure 97

PCT/US05/27239 248/487

Figure 98

**LEGEND:****M6 tot:** total extract (M6)**M6 surf prot:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I α-#:** immune sera against #**P α-#:** pre-immune sera

PCT/US2005/027239 249/487

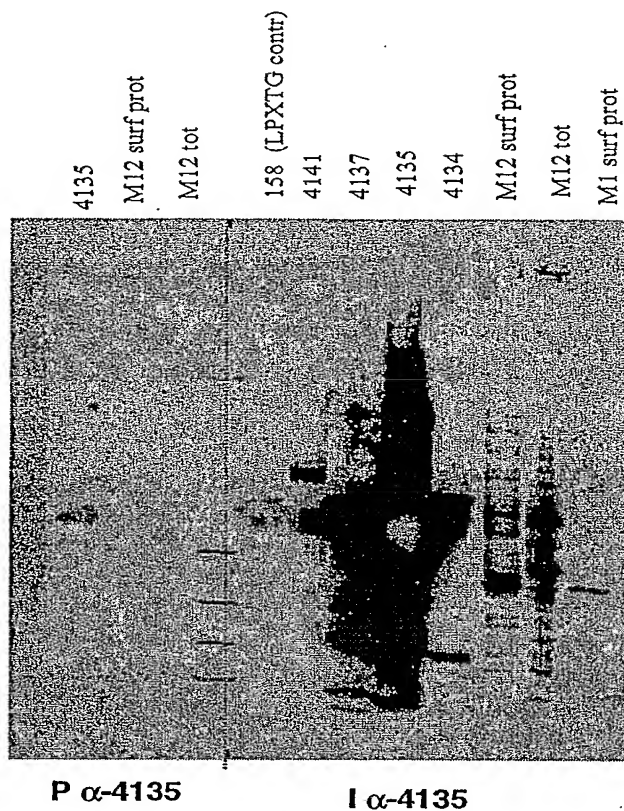
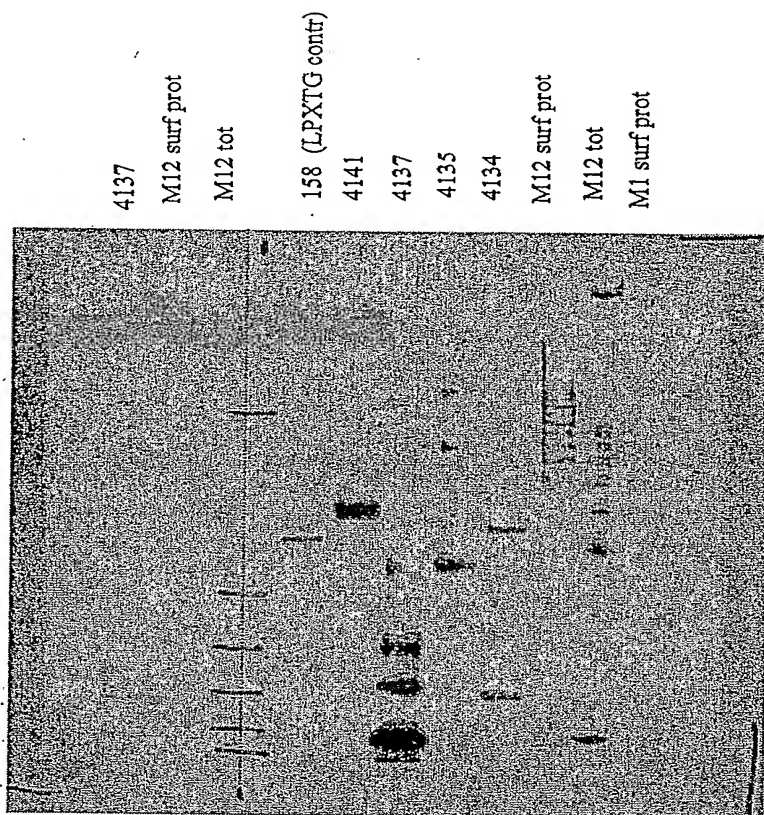
**LEGEND:****M12 tot:** total extract (M12)**M12 surf prot.:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I  $\alpha$ -#:** immune sera against #**P  $\alpha$ -#:** pre-immune sera

Figure 99

PCT/US05/27239 250/487

**LEGEND:**

M12 tot: total extract (M12)

M12 surf prot.: fraction enriched in surface proteins

#: Purified recombinant proteins, 30 ng

I α-#: immune sera against #

P α-#: pre-immune sera

P α-4137

I α-4137

Figure 100

FIGURE 101

1 GACAGCTTCCTTATACGACCGCTTCTATATCGGACCTTCCAGTTCTTGCTCTTTTA 60  
CTGTTGGAAGGAATATGCTGGCGAAAGATATAGCCTGAGAAGTTCAAGAACGAGAAAAAT  
61 CCAGGACTAGCCGTTTCAGGTGCAACGATTGTGGTGGTTGTTAAATGGAACCAAGTCGT  
GGTCCCTGATCGGCAAGTCCACGTTGCTAACAGCCACCAAAACAATTACCTTGGTCAGCA 120  
121 TCAGTTGTGACAGAAATTTACCTTCTATCTTGGGATTCCTCGTTATGTTGGAGCTAGTGCC  
AGTCAACACTGTCTTAAATGGAAGATAGAACCCCTAAGGCAATACAAACCTCGATCACGG 180  
181 TTAAGATTTTCAAAATTTGTGAAAGCCGGAGAACTCTTGAGCTTTGGGCAATGTTTTTG  
AATTCTAAAAGTTTAAACACTTTCGGCCCTCTTGAGAACTCGAAACCCGTTAACAAAAAC 240  
241 CTCCTGGTCGCGATGGGAGTAGCTTTTGGCGGTCAGATGGTGGCTATTCTGCTCTTGACC  
GAGAACCAAGCGCTACCCCTCATCGAAACAGCCAGTCGTACCAACCGATAAGCGAAGACTGG 300  
301 AGCTATGTGAAAAAACACGACTTCACCCCTTTTGGTAAATAACCGTATCGTGTGGTAGT  
TCGATACACTTTTGTGCTGAAGTGGGAAAAAACCATTTATGGCATAGCACGAACCATCA 360  
361 GTTTTGCTACTTTACAGTTTGTCCGTTTATTTGTATAAGAAAAACCTTGAGGGGTAAAC  
CAAAACGATGAATGTCAAAACAGGCAATAAACAATTTCTTTTGGAACTTCCCCCATTTG 420  
421 TCCTCAAGGTTTTATACCTTAGAAATCTCTTCAAAACCGGTCAGCTTTATCTGCAACC  
AGAAGTTCCAAATATGAGAATCTTTTAGAGAAGTTTGGCGGCAATAGACCGTTGG 480

Figure 101A

481 TCAGAAACAGTGTGAGAGCCCTGCGGCTAGCTTCCTAGTTGCTCTCTTCATTTTCATT  
-----+-----+-----+-----+-----+-----+-----+-----+  
AGTTTGTGCACAAACTCTGTCGGACGCGGATCGAAGGATCAACGAGAAACTAAAGTAA 540  
-----+-----+-----+-----+-----+-----+-----+-----+  
GAGCTTTAAATCCAGTCAGGGTAATCCCAATAGGGGACACCTCTTTCTTTCTCTCGCTT 600  
-----+-----+-----+-----+-----+-----+-----+-----+  
CTCGAAATTTTAGGTCAGTCCCATTTAGGGTTATCCGCCCTGTGGAGAAAGAAAGAGCGAA  
-----+-----+-----+-----+-----+-----+-----+-----+  
601 AATCTTCATAGAGTTGCAGGGCTATTTGGCTTATCTGACTAGCATCTTGTTTGTGTTTGG  
-----+-----+-----+-----+-----+-----+-----+-----+  
TTAAGAAGTATCTCAACGTCGCCGATAACCGAATAGACTGATCGTAGAACACAAAAACC  
-----+-----+-----+-----+-----+-----+-----+-----+  
661 CAAGACTTTTTCGTTTGGTAAGAGTTGAAAAAGTCCCTGCTAGCGGATTTTCAAAATGACAA  
-----+-----+-----+-----+-----+-----+-----+-----+  
GTTCTGAAAAAGCAACCATTTCTCACTTTTTCAGGACATCGCCTAAAGTTTACTGTT  
-----+-----+-----+-----+-----+-----+-----+-----+  
721 TTTTTCAGCTTTTCTCTGTTGATGTAGATTGAGAGGACTTTTCTGTATAGAAAGATCA  
-----+-----+-----+-----+-----+-----+-----+-----+  
AAAAAGTTCGAAAAAGAACAACTACATCTAACTCTCGCTGAAAAAGACTATCTTCTCAGT  
-----+-----+-----+-----+-----+-----+-----+-----+  
781 GCTCTTTTGTGATATCTTCTCGGACGGAGAAATCTTCCCGTAGGTTTCTCTCTTCCCGA  
-----+-----+-----+-----+-----+-----+-----+-----+  
CGAGAAAAAATATAGAGGAGCCGTGCTCTTTAGAGGGCATCCAAAAAGAGGAACGGCT  
-----+-----+-----+-----+-----+-----+-----+-----+  
841 TTGATTACGGATGCGATTGGATTGACTGGAGAGTTGTGAATGCCACGAGCCTTTCCGAT  
-----+-----+-----+-----+-----+-----+-----+-----+  
AACTAAATGCTACGCTAACCTAACTGACCTCTCAACACTTACGGTGTCTGGAAAGCTA 900  
-----+-----+-----+-----+-----+-----+-----+-----+  
961 ACAGATCATAGCCTAGCTACCAAAACGGTCTATTAGGGTTACTCAGGAACCTTCAAGTA  
-----+-----+-----+-----+-----+-----+-----+-----+  
TGTTCTAGTATCGGATCAGATGGTTTTCGACAGATAATCCCAATGGAGTCCCTTGAAGTTCTAT

253/487

Figure 101B

961 AATCAGCACCAAGTAAAAACGCCCATTTGATGAAGACGTTCTACTGCTCTTTTCCCTACTC  
-----+-----+-----+-----+-----+-----+-----+-----+  
TTAGTCGTGGTCAATTTTTCGGGTAAACTACTTCTGCAAGATGACAGAAAAAAGGATGAG  
-----+-----+-----+-----+-----+-----+-----+-----+ 1020

1021 CATGAAATTTGGAATATCCATTTGTTGAGAAAAATCCTCAGCCTGTTTCAGGTAGAATCA  
-----+-----+-----+-----+-----+-----+-----+-----+ 1080  
GTACTTTAAACCTTTATAGGTAAACAAACTCTTTTAGGAGTCGGACAAGTCCGATCTTAGT  
-----+-----+-----+-----+-----+-----+-----+-----+

1081 CTGTCAAAACCATGTGGTTTTTGATAATCACTCGCCCATTTTAGCTAAGAAATTTGTGTAAAG  
-----+-----+-----+-----+-----+-----+-----+-----+ 1140  
GACAGTTTGGTACACCAAAAACTATTAGTGAGCGGTAAAAATCGATTCTTTAAACAACATTC  
-----+-----+-----+-----+-----+-----+-----+-----+

1141 AAACGCCCTGCGAAGCAGTTAGATGGAGTCTTTCCAGATATCTTTTGAATGAGGCGAG  
-----+-----+-----+-----+-----+-----+-----+-----+ 1200  
TTTCCGGACGCCCTTCGTCAATCTACCTCAAGAAAGGCTATAGAAAAAACTTACTCCGCTC  
-----+-----+-----+-----+-----+-----+-----+-----+

1201 CAATTTGACCGCTGACTTGATACCGAGTTATTTCTGTCTACATCCAAATAGGCTTCGT  
-----+-----+-----+-----+-----+-----+-----+-----+ 1260  
GTTAAAACTGGCGACTGAACATATGGCTCAAAATAAAGACAGTGTAGGTTTATCCGAAGCA  
-----+-----+-----+-----+-----+-----+-----+-----+

1261 CAATGCTCATGGGTTCAATCAAAATCTGTATAGCGCTTAAAAATAGCTCGAATCCGGAGTC  
-----+-----+-----+-----+-----+-----+-----+-----+ 1320  
GTTACGAGTACCCAAAGTTAGTTTAGACATATCCGGAATTTTATCGAGCTTAGGCCCTCAG  
-----+-----+-----+-----+-----+-----+-----+-----+

1321 CCACAGACTTGATTTCTCATAATTCCTGAGATAAAGACAGCCCTGGGACACAGTTTCAT  
-----+-----+-----+-----+-----+-----+-----+-----+ 1380  
GGTGTCTGAACATAAAGAGTATTAAGGACTCTATTCTGTCTCGGACCCCTGTGTGCAAGTA  
-----+-----+-----+-----+-----+-----+-----+-----+

1381 AAGCTTCCTTGGAACTCATGGCAGAAATGGACACCAAAAGCTCTTTCCTCATTAACACTACAG  
-----+-----+-----+-----+-----+-----+-----+-----+ 1440  
TTCCGAAGGAACCTTGAGTACCGTCTTACCTGTGGTTTTTCGAGACGGAGTATTGATGTCC  
-----+-----+-----+-----+-----+-----+-----+-----+



PCT/US05/27239

254/487

Figure 101C

1441 TAGAAGGACTCCCGCTCCACCTGTTTCCGAGGGTCGCTTCCAATAATGACAGGTTTTTC  
-----+-----+-----+-----+-----+-----+-----+  
1500 ATCTTTGCTGAGGGCAGGTGGACAAACGGCTCCAGCGAAGGTTATTTACTGTCCAAAAG  
-----+-----+-----+-----+-----+-----+-----+  
1501 CTCTGAGTTTAGGATTATCCCTGATTCCACTGCAGCAAAAAGGCATCCATGTCATAT  
-----+-----+-----+-----+-----+-----+-----+  
1560 GAGACTCAAATCCTAATAGGGACTAAAGGTGACGTCGTTTTTTTCCGTAGGTACAGTTATA  
-----+-----+-----+-----+-----+-----+-----+  
1561 GGATGATTTTCTCTGACAAATCATTTAACAAAGGAAAAATCAACATGCCCTAGCACCTTTT  
-----+-----+-----+-----+-----+-----+-----+  
1620 CCTACTAAAAGAACTGTTTAGTAAATTGTTTCCCTTTTAGTTGTACGGATCCTGGAAAA  
-----+-----+-----+-----+-----+-----+-----+  
1621 TATACTCTTCGAAAAATCTCTTCAAAACCAAGTCAGCTTCCATCTGCAACCTCAAAACAGTA  
-----+-----+-----+-----+-----+-----+-----+  
1680 ATATGAGAAGCTTTTAGAGAAGTTTGGTGCAGTCGAGGTAGACGTTGGAGTTTGTGCTAT  
-----+-----+-----+-----+-----+-----+-----+  
1681 TTTTGAGCTGACTCGTCACTTCTATTATACAACTCAAGCAGTCTTTTGACGAGCCTGC  
-----+-----+-----+-----+-----+-----+-----+  
1740 AAAACTCGACTGAAGCAGTCAAGATAAATGTTGGAGTTTCGTACAGAACTCGTCGGAGG  
-----+-----+-----+-----+-----+-----+-----+  
1741 GGCTAGTTTCTAGTTTGCTTTTTCGATTTCCTCATTTGAGTGTAACCTGCTTATTTTCTTTTAT  
-----+-----+-----+-----+-----+-----+-----+  
1800 CCGATCAAAGGATCAAACGAAAAGCTAAAGGTAACACATTCAGCAATATAAGAAAAATA  
-----+-----+-----+-----+-----+-----+-----+  
1801 TATACCTTTTCTCGAAAAAAGAAAAAGGACTTTATTTTTCATAAATAATATAATACA  
-----+-----+-----+-----+-----+-----+-----+  
1860 ATATGGGAAAAAAGACTTTTTCCTTTTCTCTGAAATAAAAAAAGTTTATATATATGT  
-----+-----+-----+-----+-----+-----+-----+  
1861 GTTTGAAATAAATAATAGACTGTTTGTAGAAAAAGTGTAAAAATAGGAATTTTTCACCT  
-----+-----+-----+-----+-----+-----+-----+  
1920 CAAACTTATTTTATATCTGACAAATCTTTTCTTTTCTTTCACATTTTATCCTTAAAAAGTGA  
-----+-----+-----+-----+-----+-----+-----+

Figure 101D

1921 TGTGAAATCGGTACTTTATGGTATACCTGTCTCATGAATGTAACAGATGACTGTTACT + 1980  
-----+-----+-----+-----+-----+-----+-----+-----+  
ACAACTTTAGCCAAATGAATACCATATGAACAGAGTACTTACATTGTCTACTGACAAATGA  
-----+-----+-----+-----+-----+-----+-----+-----+  
AGAAAAAGAGGACATTAATATGTTGTTAAGACAGTTGTTGAAGCACAAAGATATTTTG + 2040  
-----+-----+-----+-----+-----+-----+-----+-----+  
TCCTTTTTCCTCTGTAATTATACCAACAATCTGTCAACAACACTGCTGTTCTATATAAAAC  
-----+-----+-----+-----+-----+-----+-----+-----+  
M V V K T V V E A Q D I F D - this orf is the homologue of sp0459, a formate  
acetyltransferase (pfl). It is out of the  
pilus locus  
2041 ACAAGCTTGGGAAGGCTTCAAGCGTAGATGGAAGAAAAAGCAAGTATCACGCT + 2100  
-----+-----+-----+-----+-----+-----+-----+-----+  
TGTTTCGAACCCCTTCCGAAGTTTCCGCATCTAACCTTCTTTTTCGTTTCACATAGTGCGA  
-----+-----+-----+-----+-----+-----+-----+-----+  
K A W E G F K G V D W K E K A S V S R F -  
2101 TTGTACAAGCTAACTACACACCTTATGATGGAGACGAAAGCTTCTTGCAGGACCAACAG + 2160  
-----+-----+-----+-----+-----+-----+-----+-----+  
AACATGTTCCGATGATGTGTGGAATACTACCTCTGCTTTCGAGGAACGCTGCTGTTGTC  
-----+-----+-----+-----+-----+-----+-----+-----+  
V Q A N Y T P Y D G D E S F L A G P T E -  
2161 ACGGTTCACTTCACATCAAGAAAATTGTAGAAGAACTAAAGCACACTAGCAAGAAATC + 2220  
-----+-----+-----+-----+-----+-----+-----+-----+  
TCGCAAGTGAAGTGTAGTCTTTTAAACATCTTCTTTGATTTCTGCTGTGATGCTTCTTTGAG  
-----+-----+-----+-----+-----+-----+-----+-----+  
R S L H I K K I V E E T K A H Y E E T R -  
2221 GTTTCCTCAATGGACACTCGTCCAACATCTATCGCTGATATCCCTGCTGGATTTATCGACA + 2280  
-----+-----+-----+-----+-----+-----+-----+-----+  
CAAAGGTTACCTGTGACAGGTTGTAGATAGCGACTATAGGACGACCTAAATAGCTGT  
-----+-----+-----+-----+-----+-----+-----+-----+  
F P M D T R P T S I A D I P A G F I D K -  
2281 AAGAAAATGAAGTTATCTTTGGTATCCAAAATGATGAACCTCTTCAAAATGGAACCTTCATGC + 2340  
-----+-----+-----+-----+-----+-----+-----+-----+  
TTCCTTTACTTCAATAGAAACCATAGGTTTACTACTTGAAGATTAACTTGAAGTACG

Figure 101E

C E N E V I F G I Q N D E L F K L N F M P -  
2341 CAAAAGGTGGTATCCGTATGGCTGAAACTACTTTAAAGAAATGGATACGAACCCAGACC  
-----+-----+ 2400  
GTTTTCCACCATAGGCATACCGACTTTGATGAAATTTCTTTTACCTATGCTTGGTCTGG  
-----+-----+  
C K G G I R M A E T T L K E N G Y E P D P -  
2401 CAGCTGTTACGAAATCTTCACTAAATATGTAACAACAGTTAACGACGGTATTTCCGTG  
-----+-----+ 2450  
GTCGACAAGTCTTTAGAAAGTGATTTATACATTTCTGTGTCATTGCTGCCATRAAAGGCAC  
-----+-----+  
C A V H E I F T K Y V T T V N D G I F R A -  
2461 CCTACACTCAAAATATTCGTGGGCTCGTCTCATGGCACACACTGTAACTGGTCTTCCAGATG  
-----+-----+ 2520  
GGATGTGAAGTTTATAAGCAGCGCCGACAGTACGCTGTGTGACATTGACCAGAAAGGCTTAC  
-----+-----+  
C Y T S N I R R A R H A H T V T G L P D A -  
2521 CATACTACCGGACGTATCATCGGTGTTTACGACGCTTCTCTCTTTACGGTGCAGACT  
-----+-----+ 2580  
GTATGAGTGGCCTGCATAGTAGCCACAATGCGTGCAGAACGAGAAATGCCACGCTGTA  
-----+-----+  
C Y S R G R I I G V Y A R L A L Y G A D Y -  
2581 ACTTGATGAAGAAAGTAACGACTGGAATGCAATCAAGAAATCGATGAAGAAACAA  
-----+-----+ 2640  
TGAACACTGCTCTTTTTCATTGCTGACCTTACCTTAGTTCTTTTAGCTACTTCTTTGTT  
-----+-----+  
C L M Q E K V N D W N A I K E I D E E T I -  
2641 TCCGTCTTCGTGAAGAAAGTAACCTTCAATACCAAGCATTGCAACAAGTTGTTCCGCTGG  
-----+-----+ 2700  
AGGCAGAACCACTCTTTCATTGGAAGTTATGGTTCGTACCTTGTTCACAAAGCGGACC  
-----+-----+  
C R L R E E V N L Q Y Q A L Q Q V V R L G -  
2701 GTGACCTTTACGGGTTGATGTTCCGCAAAACCAGGATGAACGTGAAGAGCAATCCAAAT  
-----+-----+ 2760

257/487

Figure 101F

CACTGGAAATGCCCAACTACAAGCGTTTGGTCGCTACTTGCACCTTCTTCGTTAGGTTA  
D L Y G V D V R K P A M N V K E A I Q W -  
GGGTTAACATTGCTTTTCATGGCTGTCTGCGGTGTGATTAAACGGTGTCTACATCTCTAG + 2820  
-----+-----+-----+-----+-----+-----+  
CCCAATTGTAACGAAAGTACCGACAGACGGCACACACTAATGCCACACGATGTAGAGATC  
V N I A F M A V C R V I N G A A T S L G -  
GTGCTGTACCAATCGTATTGGACATCTTTGACAGAACGTGACCTTCTCGTGGTACATTTA + 2880  
-----+-----+-----+-----+-----+-----+  
CAGCACATGGTAGCATTAACCTGTAGAAAGCTCTTGCACTGGAACGAGCACCATGTAAT  
R V P I V L D I F A E R D L A R G T F T -  
CTGAATCAGAAATCCAAGAAATTCGTTGATGATTTCGTTATGAAACTTCGTACAGTTAAAT + 2940  
-----+-----+-----+-----+-----+-----+  
GACTTAGTCTTTAGGTTCTTAAGCAACTACTAAAGCAATACTTTGAAGCATCTCAATTTA  
E S E I Q E F V D D F V M K L R T V K F -  
TTGCTCGTACCAAGCTTATGACCAATTGTTACTACTAGTGACCCACCTTTATCACAACCTT + 3000  
-----+-----+-----+-----+-----+-----+  
AACGAGCATGGTTTCGAATACTGGTTAACATGAGTCCACTGGGTGGAAATAGTGTGAA  
A R T K A Y D Q L Y S G D P T F I T T S -  
CTATGGCTGGTATGGGTAAACGACGGTGTGTCACCGTGTACTAAGATGGACTACCGTTTCT + 3060  
-----+-----+-----+-----+-----+-----+  
GATACCGACCATACCCATTGCTGCCAGCAGTGGCACAATGATTCTACTGATGGCAAGA  
M A G M G N D G R H R V T K M D Y R F L -  
TGACACTCTTGACAACTCGGTAACTCACCAGAACCAAACTTGACAGTCTTTGGACTG + 3120  
-----+-----+-----+-----+-----+-----+  
ACTTGTGAGAACTGTTGTAGCCATTGAGTGGTCTTGGTTTGAAGTGTCAAGAAACCTGAC  
N T L D N I G N S P E P N L T V L W T D -

PCT/US05/27239 258/487

Figure 101G

3121 ~ACAAATGCCATACAACTCCGTCGCTACTGTATGCACATGAGCCACAAAACACTCTTCTA 3180  
+-----+-----+-----+-----+-----+-----+  
TGTTTAAGGATATGTTGAAGGACGCGATGACATACGTTACTCGGTGTTGTGAGAAGAT  
+-----+-----+-----+-----+-----+-----+  
K L P Y N F R R Y C M H M S H K H S S I -  
+-----+-----+-----+-----+-----+-----+  
TCCAAACGAAGGTGTAAACAACATGGCTAAAGACGGATATGGTGAATGAGCTGTATCT 3240  
+-----+-----+-----+-----+-----+-----+  
AGGTTATGCTTCCACATGTTGTTACCGATTCTCTGCCCTATACCACCTTTACTCGACATAGA  
+-----+-----+-----+-----+-----+-----+  
Q Y E G V T T M A K D G Y G E M S C I S -  
+-----+-----+-----+-----+-----+-----+  
CATGCTGTGTCTCCACTTGATCCAGAAAATGAAGAACAACGCCACACATCCAGTACT 3300  
+-----+-----+-----+-----+-----+-----+  
GTACGACACACAGAGGTGAACGTAGGCTCTTTACTTCTTGTGCGGTGTAGGTCATGA  
+-----+-----+-----+-----+-----+-----+  
C C V S P L D P E N E E Q R H N I Q Y F -  
+-----+-----+-----+-----+-----+-----+  
TCGGTCTCGTGTAAACGTTCTTAAAGCCCTTCTTACTGTTTGAATGGTGGTTACGACG 3360  
+-----+-----+-----+-----+-----+-----+  
AGCCACGAGCACATTTGCAAGATTTCGGGAAGAAATGACCAAACTTACCACCAATGCTGC  
+-----+-----+-----+-----+-----+-----+  
G A R V N V L K A L L T G L N G G Y D D -  
+-----+-----+-----+-----+-----+-----+  
ATGTTACAAAGACTACAAAGTATTTGATATCGAACCATCCGTGACGAAGTTCTTGAAT 3420  
+-----+-----+-----+-----+-----+-----+  
TACAAAGTGTTCATGATGTTTCATAAACTATAGCTTGGTTAGGCACCTGCTTCAAGAACTTA  
+-----+-----+-----+-----+-----+-----+  
V H K D Y K V F D I E P I R D E V L E F -  
+-----+-----+-----+-----+-----+-----+  
TTGAATCAGTTAAAGCGAACTTTGAAAATCTCTTGAAGTGGTTGACTGACACTACGTTAG 3480  
+-----+-----+-----+-----+-----+-----+  
AACTTAGTCAATTCGCTTGAACCTTTTATAGAGAACTGACCAACTGACTGTGAATGCATC  
+-----+-----+-----+-----+-----+-----+  
E S V K A N F E K S L D W L T D T Y V D -  
+-----+-----+-----+-----+-----+-----+  
ATGCTTGAACATCATCCACTACATGATAGGTACAACTACGAAGCTGTTCAAATGG 3540  
+-----+-----+-----+-----+-----+-----+  
TACGGAACCTGTAGTAGGTGATGTACTGACTATCCATGTTGATGCTTCGACAAAGTTTACC

PCT/US05/27239 259/487

Figure 101H

C A L N I I H Y M T D R Y N Y E A V Q M A -  
3541 CCTTCTTGCCAACTAAACAACGGTGCCCAACATGGGATTCGGTATCTGTGGATTGCTTAACA + 3600  
GGAAGAACGGTTGATTTGTTGCACGGTTGTACCCCTAAGCCATAGACACCTAAACGATGTG  
C F L P T K Q R A N M G F G I C G F A N T -  
3601 CTGTTGATACATTGTCAGCTATCAAAATACGCTACAGTTAAACCAATCCGCTGACGAAGATG + 3660  
GACAACTATGTAACAGTCGATAGTTTATGCGATGTCAATTTGGTTAGGCACCTGCTTCTAC  
C V D T L S A I K Y A T V K P I R D E D G -  
3661 GCTACATCTACGATTACGAACAATCGGTGACTACCCACGCTGGGGTGAAGATGACCCAC + 3720  
CGATCTAGATGCTAATGCTTTGTTAGCCACTGATGGTGGCGACCCACTTCTACTGGGTG  
C Y I Y D Y E T I G D Y P R W G E D D P R -  
3721 GTTCAACGAATTGGCAGAAATGGTTGATCGAAGCTTACACAACCTCTCTACGTAGCCACA + 3780  
CAAGTTGCTTAACCGTCTTTACCACTAGCTTCGAATGTGTGAGCAGATGCATCGGTGT  
C S N E L A E W L I E A Y T T R L R S H K -  
3781 AACATACAAAGACGCGAGAGCTAGTATCCTTTTGACAATCACAATCACTAAGTTGCTT + 3840  
TTGATATGTTTCTGCGTCTTCGATGTCTAGTGTAAGTGAAGTGTAGTAGATTCCAAACGAA  
C L Y K D A E A T V S L L T I T S N V A Y -  
3841 ACTCTAAACAACTGGTAACCTACCAAGTTCACAAAAGTGTATACCTCAACGAAGATGGTT + 3900  
TGAGATTGTTTTCACCATTTGAGTGGTCAAGTGTTCACATATGGAGTTGCTTCTTACCAA  
C S K Q T G N S P V H K G V Y L N E D G S -  
3901 CTGTGAACCTTGCTAACTTGAATCTTCTCACCAGTGTCAACCATCTAACAAGCTA + 3960

PCT/US05/27239 260/487

Figure 101I

GACACTTGAACAGATTGAACTTAAGAAGAGTGGTCCACGATGGGTAGATTGTTTCGAT  
V N L S K L E F S P G A N P S N K A K -  
AAGTGGTGGTTCGAAAACTTGAACCTCAGTCTTCTAGCCCTTGACTTTAGTTATGCAGCTG + 3961  
TTCCACCAACCAACGTTTGAACCTTGAGTGAAGATCGGAAGTGAATCAATACGTCGAC  
G G W L Q N L N S L S L D F S Y A A D -  
ACGGTATCTCATTGACTACACAAAGTATCACCTCGCGCTCTTGGTAAGACTCGTGATGAAC + 4080  
TGCCATAGAGTAACCTGATGTGTTTCATAGTGGAGCGCGAGAACCATTCCTGAGCACACTTGG  
G I S L T T Q V S P R A L G K T R D E Q -  
AAGTGTATACTTGGTAACAATCCTTGATGGTTACTTCGAAAACGGTGGACAAACAGCTTA + 4081  
TTCAACTANTGAACCATGTTAGGAACCTAGCAACTACCAATGAAGCTTTTGCCACCTGTTGTGCAAT  
V D N L V T I L D G Y F E N G G Q H V N -  
ACTTGAACGTTATGGACTTGAACGATGTTTACGAAAAAATCATGTCAGGCGAAGACGTTA + 4141  
TGAACCTGCAATACCTGAACCTTGCTACAAAATGCTTTTATAGTACAGTCCGCTTCTGCAAT  
L N V M D L N D V Y E K I M S G E D V I -  
TCGTACGTATCTCTGGATACCTGTAAACACATAAATACCTCCTCCTCAGAACAAAAAATCG + 4260  
AGCATGCATAGAGACCTATGACACATTTGTGATTTATGGAGTGAAGTCTGTTTGTGAC  
V R I S G Y C V N T K Y L T P E Q K T E -  
AATTGACACAACGCTGCTTCCACGAAGTCTTCAATGGATGACGCCCTTGGATGCAATTGA + 4261  
TTAAGTGTGTTGCACAGAAGGTGCTTCAAGAAAGTTACCTACTGCGGAACCTACGTAACCT  
L T Q R V F H E V L S M D D A L D A L S -



PCT/US05/27239

261/487

Figure 101J

4321 GCTAATCAAGTCTTGAATAATAAAGGGCTCTTTGTCAACTGTAGTGGTGAAGAAA 4380  
-----+-----+-----+-----+-----+-----+  
CGATTAGTTCAGAAGCTTATTTATTTTCCCGAGAAACAGTTGACATCACCCAACTCTTTT  
c \*  
4381 AGCTAAGCTCGAGAAAGGACAAATTTGTCCCTTCTTTTGTATGTTTCAGAGCGATGAAA 4440  
-----+-----+-----+-----+-----+-----+  
TCGATTCGAGCTCTTTCCTGTTTAAACAGGAAAGAAAAAACTACAAGTCTCGCTACTTTT  
e \* A R S L V F K T R E K K I N L A I F -orf1\_670 homologue of sp0460, transposase  
4441 ATCCGTTTTTTGAAGTTTCAAAAGTTCCGAAACCCAAAGGCATTGCGCTTGATGCTTTTG 4500  
-----+-----+-----+-----+-----+-----+  
TAGGCAAAAACCTCAAAAGTTTCAAGGCTTTTGGTTTCCGTAACGCGAAGTACAGAAAC  
e I R K K F N E F N R F G F A N R K I D K -  
4501 ATGAGTTTGTAGTGGCTCAAGTTTAGCGTTAGATAAAGGCAATTCAATGGCGTTAGTG 4560  
-----+-----+-----+-----+-----+-----+  
TACTCAACAAATCACGGAGTTCAATCGCAATCTTATTCGGTTAAGTTACCGCAATCAC  
e I L K N T A E L K A N S Y P L E I A N T -  
4561 ATGTAGTTTTTATAGCAATAAATGTGCTCAAAGTGGTTTAAAGGTGCGGTTGAGATGA 4620  
-----+-----+-----+-----+-----+-----+  
TACATCAAAAATATCGTTTATTTACACGAGTTTCACCAAAATTTCCACGCCAACTCTACT  
e I Y N K Y C I F T S L T T K F T R N L H -  
4621 GGTACGTGCTTGAATTAAGCCCCCAAACTGGTCAGTATTCCTCTTGTAGATGAAAT 4680  
-----+-----+-----+-----+-----+-----+  
CCATTGCACAGAACTTAATTCGGGGTTTGACCACTCATAGAAGAGACATCTACTTTA  
e P L T D Q I L G W F Q D T N K E Q L H F -  
4681 AGGAGTAGTTGATACAGTCAATGTAATCTTTAAGTTCAGGTACTAGAGTAAAGATTTTC 4740  
-----+-----+-----+-----+-----+-----+  
TCCTCATCAACTATGTCAGTATCATATAGAAAATTCAGTCCATGATCTCATTTCTAAAG

262/487

Figure 101K

e L L L Q Y L D Y Y D K L E P V L T F I K -  
4741 TTGAGACACTCCCTAGGAGTTAAGGTCTCTGTGAAAGTTCTAGCATAGAAAGGCTTAAGA + 4800  
AAGTCTGTGAGGATCCCTCAATTCAGAGACATTCAGATCGTATCTTCCGAATTC  
K L C E R P T L T E R F T R A Y F P K L -  
4801 GAGAGTTCCGACTATCTTTTAGGATAAATTTCCAGTAATATTTAAGAGCCTGTATTCC + 4860  
CTCTCAAAGGCTGATAGAAAATCCCTATTAAAGGTCANTATATAAATTCGAGACATAAGG  
S L K R S D K L I F K W Y Y K L A R Y E -  
4861 AGAGATTATCATCAAAATTCCTTCATGATGTTGATTCTAGTCTGATTAAGACCCCTGCTC + 4920  
TCTCTAAATAGTAGTTTAAAGGAAGTACTACAACCTAAGATCAGACATAATTCGCGGACGAG  
L S K D D F Q K M I N I R T Q N L A R S -  
4921 ATGTGTTGGACAATGTGGAACGATCGAGAACAATTTAGCATGGGAAATAATTTCTTA + 4980  
TACACAACCTGTTACACCTTTGCTAGCTCTGTTAAATCGTAACCTTTTATTAAGAAT  
M H Q V I H F R D L V I K A N P F L K K -  
4981 ATGAGAGGGATATAACTTCCAGACATATCAACAGTGACGACTTAACTTTTCTAGCT + 5040  
TACTCTCCCTATATTGAAGGTCTGTATAGTTGTCTACTGCTGAAATGAAAAAAGATCGA  
I L P I Y S G S M D V T V V K V K K R A -  
5041 TCTTTCGAGTACTTGAGAAATGATTTCCGGATGGTTGTTTGGACGCTGTGTATCAAGAATG + 5100  
AGAAAGCTCATGAACCTTCTTACTAAAGCCCTACCAACAACCTGCAGACATAGTTCTTAC  
E K S Y K F F H N R I T T Q R R N D L I -  
5101 GTCATGATTTTCTTAGTGTGAAATCCTGAGCAATGAAGCAATTTCCCTTCTGCTAG + 5160

### Figure 101b

CAGTACTTAAAGAATCACAACTTTAGGACTCGTTACTTTCCGTTAAGGGGAAGACCATC

e  
F  
M  
I  
K  
K  
T  
N  
F  
D  
Q  
A  
I  
F  
A  
L  
K  
G  
K  
O  
Y

5161 GAGAAATTCATCCCGAGAGAGGATTTTCAGGCCAAAGTGGTGTAATCTCTTGGAAATGGAAT  
5220 CTCTTAAGTAGGGTCCCTCTCTCCTAAAGTCCGTTTACCACCATTTAGGAGAACCTTTACTTTA

e S F E D W S L I E P L T T Y D E Q F H F -

TGCTTGAGCTTACGATAGACGGTAGAGGTAGAGATGGCTAATTTAGAGCGGATA  
 5221  
 -----+-----+-----+-----+-----+-----+  
 ACGAATCGGAATGCGATCTGCCATCTCCATCTCTACCGATTAAATCTTCGGTAT  
 5280

Q K L K R Y V T S T S I A L K S A I -

5281 TGTGTAAGAGCCCTCTCTGTTGAGTAGGAGTTGGGCAATTCTGTCTCACCATTTCCGAG  
-----+-----+-----+-----+-----+-----+-----+  
ACACATTCTCGGAGAGACAACCTGATCTCTCAACCCGTTAAAGACACAGTGGTAAAGGCTC  
5340

H F L A E R N L L Q A I K Q R V M E S -

5341  
ATTTGGCAATTTTTCGTGAACGAGAGTTGTTTCAGCTACAGTGAATTTCCGACAGGACTTG  
-----+-----+-----+-----+-----+-----+  
TAAACCGTTAAAAAGACTTGCTCTCAACAAACTCGATGTCACCTGAAGAGCTGTCTCTGAAC

Q C N K Q V L T F E A V T V K R C S K

5401 CATCGAAATCGTCTCTCTTTTTCAAATGAATAGGCTAGGGAACACCAATCTCGATAAAA  
-----+-----+-----+-----+-----+-----+-----+  
5460 GTAACTTTAGCAGAGAAAAGTTTACTTACTCCGATCCCTTGGTGGTTAGAGCAATTTT

e  
c  
u  
q  
r  
r  
r  
k  
k  
l  
h  
h  
l  
s  
p  
r  
g  
g  
h  
h  
h  
h  
i

5461  
GGGATTTTAGAAGGCTTTTGGAAAGTCGATATTTGATTTCTTTACAGTGTTTACAT  
-----+-----+-----+-----+-----+-----+-----+  
CCCTAAAATCTTCCGAAAAACCTTCAGCATATAACTAAACAAAAGGAATGTCACAAATGTA  
-----+-----+-----+-----+-----+-----+-----+ 5520

a  
 P I K S P K Q F D Y K I O K G K C H K C -

Figure 101M

TTAGTGGTGATATCAAGTGTAGCGAAGACTTCGATATGGGTATCGTCTGAATGGCT + 5580  
-----+-----+-----+-----+-----+-----+  
AATCCACCCACTATTAGTTCACATCGCTTCTGAAGCTATACCCATAGCAGGACTTACCGA  
K P P H Y D L T A F V E I H T D H Q I A -  
TTATTTAAGGTGATGTTTTTTTGTCTTTTATTCGATGAGTAATGTGGTATGATGTGT + 5640  
-----+-----+-----+-----+-----+-----+  
AATAAATCCCACTACAAAAACAGAAAAATAAGGCTACTCATTACACCACTAATACTACACA  
K N L T I N K D K I G I L L T T H N I H -  
TCCATAAGATACTTTTCTAATGAGTTGTTTGTAGGGCTTTTTCATTATATAAGTCTTATGGGACT + 5700  
-----+-----+-----+-----+-----+-----+  
AGGTATTCATGAAAGATTACTCAACAAATCCCGGAAAGTAATATTCAGAAATACCCCTGA  
E M  
TTTTTGATACTCAAAAAGCCCTATAAATCTCCACAGTGGGATTTACCCACTACAGAAATTA + 5760  
-----+-----+-----+-----+-----+-----+  
AAAAACTATGAGTTTTTTCGGGATATTAGAGGTGTCAACCCCTAAATGGGTCATGCTTTAAT  
TAGAGCCAGAAAAAACACTTTTGTCTACTAGCAGAAACTAGAGAGCAGAGTGTTTTCT + 5820  
-----+-----+-----+-----+-----+-----+  
ATCTCGGTCCTTTTGTGAAAAACAAGTGTATCGTCTTTGATCTCTCGTCTTCACAAAAAGA  
GTTCAGATTTACCCAAAACTGGGAAATATGGGGATAAGAAATAGAGATGGCTTAGGAAGCC + 5880  
-----+-----+-----+-----+-----+-----+  
CAAGTCTAAATGGGTTTGTACCCCTTATATACCCCTATCTTATCTCTACCGAATCCTTCGG  
CCTTTTGTGTGTAGACAGTACGATGAACCTTATAACAAATAGTGAGCCCTTTTATGCAATC + 5940  
-----+-----+-----+-----+-----+-----+  
GGAAAAACACAGATCTGTCTCATGCTACTTGAATATTGTTTATCTACTCGGAAAAAATCGTTAG  
\* L L Y H A K K A I -orf2\_670 homologue of sp0461, transcriptional  
regulator  
ATTGCCACCCGTTTGTCAAAAGCCCTCTTTTTCGGATATCTACAATGTGTGATAGATGAGA

$$265/487$$
[illegible]

266/487

Figure 1010

```
6361 TCCTGAAACAATTGCTTTTCGAAATATGATACAGTGGCTTGTGCTTCAATCCCATTAAGT + 6420
-----+-----+-----+-----+-----+-----+-----+
AGAACTTGTTAACGAAAGCTTTATACATATGTCACCGACAGCGAAGTTAGGGTATTACA
e E Q V I A K S I H Y L P K D S E I G Y H -

6421 TCGTAATAATTATAATAGGGAACTAGATTTGTAAACCAAAACAACGTTCTTGTTAAG + 6480
-----+-----+-----+-----+-----+-----+-----+
AGCATTAATAATATTATCCCTTGATCTAAACATTTGGTTTGTGTTTTCGAAGAACAAATTC
e E Y Y N Y Y P V L N Q L G F L F T R T L -

6481 AAATCAGTGTCTTTAAAGAAAGAGAAATTCGAATGTCATTTCTTAAGATATCTTG + 6540
-----+-----+-----+-----+-----+-----+-----+
TTTCAGTCACGACAATTTTCTCTTAAGCTTTACAGTAAAGGATTTCTATAAGAAC
e F T L A T L F S L S N S I D N G L I N K -

6541 AACTGGATAGTAGATGCTTTCCTCTTGTATGCTGAAGAATCAGTGAATAGTATGATC + 6600
-----+-----+-----+-----+-----+-----+-----+
TTGAACCTATCATCTACGAAAGGAGAACATACGACTTCTTAGTCACTTATCATCTACTCAG
e F K S L L H K G R T H Q L I L Q I T H T -

6601 TTTTTCCTTGATCCATTTTGTCTTGGAAACGAGAATTAGCAGAACATAAACCAA + 6660
-----+-----+-----+-----+-----+-----+-----+
AAAAAAGAACTAAGGTAAACAGGAACCTTTTGTCTTCTTAATCGTCTTGTATTGTTT
e K K E Q N W K D K S F S S N A S C Y V L -

6661 AAGATATAATCCAGTTCCTCTGAGTAAAGTCATGTGGCATGTGGCTCTAAGTAAGTT + 6720
-----+-----+-----+-----+-----+-----+-----+
TTCATATTAGGTCAAGAAGGACTCATTTTCAGTACACCGTACACCGAGATTCATCAA
e F I Y D L E E Q T F T M N A H P E L Y T -

6721 TGGCAATGTTCCATCAAAATCGGATACATAAAGAGTTTATTTTCAAACTCTTTG + 6780
-----+-----+-----+-----+-----+-----+-----+
ACCTTACAAGTAGTTTTCCTATGTATTCTCCAAAAAATTAAGTTTGAGAAC
```

e	Q	C	H	E	M	L	I	P	Y	M	F	L	N	K	L	K	E	F	E	K	-
	G	A	C	T	C	A	G	G	A	A	T	C	C	G	A	C	G	T	T	C	C
6781																					6840
	C	T	G	A	G	T	C	C	T	T	A	A	G	G	C	T	C	A	C	T	C
	S	E	P	F	E	L	P	F	E	R	R	K	W	T	L	A	V	L	I	S	-
	A	A	T	G	A	A	C	A	T	A	C	T	C	G	T	G	A	T	T	C	T
6841																					6900
	T	T	A	C	T	T	G	A	T	G	A	C	C	A	C	A	T	T	A	A	G
	F	H	V	Y	E	D	P	T	I	E	L	L	E	H	S	L	Q	S	N	S	-
	T	G	C	A	A	T	C	A	T	A	T	G	T	G	A	C	C	A	A	T	C
6901																					6960
	A	C	G	T	G	T	A	T	A	C	A	C	A	C	T	G	T	T	A	G	A
	Q	V	I	M	H	T	V	W	D	M	S	G	D	N	L	D	Y	I	E	I	-
	C	C	A	A	A	T	G	A	A	A	C	G	A	T	T	A	A	A	A	A	C
6961																					7020
	G	G	T	T	A	C	T	T	G	A	C	T	T	T	G	C	T	A	A	G	T
	G	F	H	F	Q	L	L	A	I	L	F	R	I	R	Y	E	P	G	V	V	-
	T	G	A	T	T	T	C	A	A	G	T	C	A	A	C	C	T	A	G	T	A
7021																					7080
	A	C	T	A	A	A	A	G	T	T	C	C	A	G	T	T	G	G	A	C	T
	Q	N	K	V	L	D	L	G	V	S	R	L	L	L	G	C	K	Q	R	V	-
	C	G	G	T	A	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	7140
7081																					7140
	G	C	A	T	C	G	A	C	A	C	T	T	A	T	A	T	A	G	A	A	A
	R	Y	G	T	A	I	S	I	Y	E	K	Q	T	F	E	N	F	S	Q	N	-
	C	C	T	T	G	T	A	G	A	A	G	C	G	A	C	T	A	T	T	T	T
7141																					7200
	C	C	T	T	G	T	A	G	A	A	G	C	G	A	C	T	A	T	T	T	T



268/487

Figure 101Q

7201  
e G Q L L F F R L I K L I T S Q N Y L Q H -  
GGAACATCATCTTCTTGGCCATATAAAATTTTATCAACTAACCAATATTTGACTACC  
AAGTAATAATTCGTTTGATGAGAAATGGTTCGATTAAATGAACCTGTTGGTATCTAAA  
TTCTATATTAAAGCAAACTACTCTTACCACAGCTAATTAACCTTGAACAACGCATAGATTT  
F Y Y N T Q H S H E I L Q V Q Q T D L -  
TTAAATGTCACACTCTCCGAAATGTTTCTGTGAATTCCTGCAAAATGCTTAGAGACTT  
AATTTACAGTTGAGAAGGAGCTTACAAAGACATTAAGGACGTTTACGAATCCTCTGAA  
N F T L E E E F T E Q L E Q L I S L S -  
TTAGATTGTAATGAAGTTAAAGTAGACAGTTCATCTAGTCAATAGACCGAATATCCAAT  
AATCTAACATTACTTCAATTTCACTCTGTCAAGTAGATCAAGTTAATCTGGCTTATAGGTTA  
K S Q L S T L T S L E D L E I S R I D L -  
AATATATTATAAATGGTAAATTTTATCTGTGAATCTTTTTCATATGATTTGTTTAGCATA  
TTATATATAAATTTTACCATTAAATATAGACATTAAGAAAAGTTACATAAACAATCGTAT  
L I N L I T I K D T I R K E I Y K N L M -  
GTACCGAATCTTAGTTGCATATAGATAATTTTAAATATTATATAACAAAAGAACTAAT  
CAATGGCTTAGAATCAACGTTATCTATTAAAAATTAATAATATTGTTTTCTTTGATTA  
TGCTCTGTCAAAAAGGTTGTGGAATTTCCGACTTTTATTGATAAAACAGCATGTAATAAAA  
ACAGAACAGTTTTCCTCCACACCTTAAAGGCTGAAATACTATTTTGTCTGATATATTTT  
GGCATTTTAAAGATAGTAATGAGTATTTGGTGGAGTTTATGCTTATTTTTTTTATTAGA  
7260  
7320  
7380  
7440  
7500  
7560  
7620

Figure 101R

CCGTAAATTTCTATCATTTACTCATTAACCCACTCAAAATACCGAATAAAAAAATAATCT  
7621  
AAATATTTTATCAAAATATTGTCGTTCTATATAAAAAATATGATGATAAAAAATATCTATT  
7680  
TTTATAAAAAAATAGTTTATTAACAGCAAGATATTTTATTTATACACTATTTTATATAGATAA  
7681  
GTGATGGAAGTGTGTTTAAATTTATCTAGGATAGTTAATAAGTAACTATCTATCTAT  
7740  
CACTACCTTCAACAAATTAATATGATCCTATCAATTTATCATTTATGATGATGATGATA  
7741  
ATTGTATACAAAGTGTGTCATGTCAGGTTGAGAGATAGCTATAACGCACCTTTTATAGGC  
7800  
TAACATATGTTACACACAGTAACGGTCCAACTCTTCTATCGATATTGCGTAAATATGCG  
7801  
TTTTGCTACGTTTGTAGTGAACGGATTAACTCAGTGAGATAAATTTTATCAGAACATAA  
7860  
AAAACGATGCAAAACAATCAGCTTGCCCTAAATGAGTGCACTCTATTAAATAGTCTTGTATT  
7861  
GTAATCCGTTTCTTCGTTGATACAGATTGAAAGTACCTATGATCATAGAGGATTAACT  
7920  
CATTAGGCAAGAAGCACATATGTCATACTTTTCATGGATACCTAGTATCTCTCTAATTGA  
7921  
TGTTCTATGAATAATGCTTAACAGGAGACACACATGAAAAAGTAAGAAGATATTCA  
7980  
ACAAGATCTTATTACGAATTGTCCTCTGTTGTTGTTTCTTTTCTTTCTATATAAGT  
7981  
GAAGGCAGTTGCAGGACTGTGCTGATATCTCAGTGACAGCTTTTCTTCGATAGTTGC  
8040  
CTTCGGTCAACGTCCTGACACGACATATAGAGTCACTGTGCGAAAAAGAGCTATCAACG  
b M L N R E T H M K K V R K I F Q -orf3\_670 homologue of sp0462, LPXTG  
b K A V A G L C C I S Q L T A F S S I V A -

```

8041      TTTAGCAGAAACGCCCTGAAACCCAGTCCACCGGATAGGAAAGTAGTGATTAAGGAGACAGG
      -----+-----+-----+-----+-----+-----+-----+
      AAATCGTCTTTGCGGACHTTTGGTCAGGTGCGCTATCCCTTTTCATCACTAATTTCTCTGTGCC
      -----+-----+-----+-----+-----+-----+-----+
      L A E T P E T S P A I G K V V I K E T G -
      -----+-----+-----+-----+-----+-----+-----+
      CGAAGGAGGAGCGCTTCTAGGAGATGCGCGTCTTTGAGTTGCAAAAAACAATACGGATGGCAC
      -----+-----+-----+-----+-----+-----+-----+
      GCTTCCTCCTCGCGAAGATCCTCTACGGCAGAACTCAACTTTTGTATTGCGCTACCGGTG
      -----+-----+-----+-----+-----+-----+-----+
      E G G A L L G D A V F E L K N N T D G T -
      -----+-----+-----+-----+-----+-----+-----+
      AACTGTTTCGAAAGGACAGAGGCGGCAACACAGGAGAAGCGATATTTTCAACATAAAACC
      -----+-----+-----+-----+-----+-----+-----+
      TTGACAAAAGCGTTTCCCTGCTCCTCGCGTTTGTCTCTCTCGCTATAAAAGTTTGTATTTTGG
      -----+-----+-----+-----+-----+-----+-----+
      T V S Q R T E A Q T G E A I F S N I K P -
      -----+-----+-----+-----+-----+-----+-----+
      TGGGACATACACCTTGACAGAAGCCCAACCTCCAGTTGGTTATATAACCCCTCTACTAAACA
      -----+-----+-----+-----+-----+-----+-----+
      ACCCTGTATGTGGAACHTGCTCTTCGGGTGGAGGTCAACCAATATTTGGGGAGATGATTTGT
      -----+-----+-----+-----+-----+-----+-----+
      G T Y T L T E A Q P P V G Y K P S T K Q -
      -----+-----+-----+-----+-----+-----+-----+
      ATGGACTGTGTAAGTTGAGAAGAAATGGTCGGACGACTGTCCAAAGTGAACAGGTAGAAAA
      -----+-----+-----+-----+-----+-----+-----+
      TACCTGCAACACTTCAACCTCTCTTACACGCTGCTGTGACAGGTTCCACTTGTCCATCTTTT
      -----+-----+-----+-----+-----+-----+-----+
      W T V E V E K N G R T T V Q G E Q V E N -
      -----+-----+-----+-----+-----+-----+-----+
      TCGAGAAGAGGCTCTATCTGACCACGATCCACAAACAGGGACTTATCCAGATGTTCAAC
      -----+-----+-----+-----+-----+-----+-----+
      AGCTCTTCTCCGAGATAGACTGGTCTATAGGTGTTTGTCCCTGAATAGGTCATACAAGTTTG
      -----+-----+-----+-----+-----+-----+-----+
      R E E A L S D Q Y P Q T G T Y P D V Q T -
      -----+-----+-----+-----+-----+-----+-----+
      ACCTTATCAGATATTATAGGTAGATGGTTCGGAAAAAAACAAGGACAGCACAGCGGTTGAA
      -----+-----+-----+-----+-----+-----+-----+
      TGGAAATAGTCTAATAATTCACATCTACCAAGCCCTTTTTCCTGCTGCTTCCGCAACTT
      -----+-----+-----+-----+-----+-----+-----+
8401      TGGAAATAGTCTAATAATTCACATCTACCAAGCCCTTTTTCCTGCTGCTTCCGCAACTT
      -----+-----+-----+-----+-----+-----+-----+

```

PCT/US2005/027239

271/487

Figure 101T

b P Y Q I I K V D G S E K N G Q H K A L N -  
TCCGAATCCATATGAACGTGTGATTCCAGAGGTACACCTTCAAGAGAAATTTATCAAGT 8520  
8461 AGGCTTAGGTATACCTGCACACTAAGGTCTTCCATGTGAAAGTTTCTCTTAATAATAGTTCA  
P N P Y E R V I P E G T L S K R I Y Q V -  
GAATAATTGGATGATAACCAATATGGAATCGAGTTGACGGTTAGTGTAATAACACGAGGT 8580  
8521 CTTATTAAACCTACTATTTGGTTATACCTTAGCTCAACTGCCAATCACCATTTTTGGCTGCCA  
N N L D D N Q Y G I E L T V S G K T T V -  
TGAAACGAAAGAAGCCTCTACTCCGCTAGATGTGTATTCTATTAGATAACTCCAATAG 8640  
8581 ACTTTGCTTTCTTCGGAGATGAGCGGATCTACAACAATAAGATAATCTATTGAGGTTATC  
E T K E A S T P L D V V I L L D N S N S -  
TATGAGTAATATTCGACATAATCATGCCCCATCGAGCGGAAAAAGCGGAGAAAGCGACACG 8700  
8641 ATACTCATATATAAGCTGTATTAGTAGGGGTAGCTCGCCCTTTTTCGCCCTCTTCCGCTGTGC  
M S N I R H N H A H R A E K A G E A T R -  
AGCCCTTGTAGATAAGATTACCTCCAATCCAGATAATCGAGTAGCCTGTGTGACTTATGG 8760  
8701 TCGGGAACATCTATTCTTAATGGAGGTTAGGTAGTCTATTAGCTCATCTGTGAACACTGAATACC  
A L V D K I T S N P D N R V A L V T Y G -  
CTCAACTATCTTTGACGGTTTCAGAACTACTGTGGAAAAAGGGGTAGCAGATCGGAACGG 8820  
8761 GAGTGTAGAGAAACCTGCCAAGTCTTCGATGACACCTTTTTCGCCCATCGTCTACGGCTGCC  
S T I F D G S E A T V E K G V A D A N G -  
AAAAATATTGAATGACTCAGCTTTATGGACGTTTCGATCGTAGACGTTTACAGCTAAAC 8880  
8821

Figure 101v

b	8881	TTTTTATAACTTACTGAGTCGAAATACCTGCAAGCTAGCATGCTGCAAAATGTCGATTTTG K I L N D S A L W T F D R T T F T A K T - TTATAATTATAGCTTTTAAATCTCAGATCAGATCCTACTGATATTCAAACTATTAAAGCA -----+-----+-----+-----+-----+-----+-----+-----+ 8940 AATATTAAATATCGGAAAAATTTAGAGTGCTAGCTAGGATGACTATAGTTTGATAATTCCT Y N Y S F L N L T S D P T D I Q T I K D - TAGGATTTCCATCAGATGCAGAGGAATTGAACAAAGACAAATTGATGATATCAATTCGGCGC -----+-----+-----+-----+-----+-----+-----+-----+ 9000 ATCCTAAGGTAGCTACGTCCTTAACTTGTTCTGTTTAACTACATAGTTAAGCGCGG R I P S D A E E L N K D K L M Y Q F G A - GACHTTTTACCCAGAGGCTTTGATGACCGCTGATGATATCTTGACAAAGCAGCAAGACC -----+-----+-----+-----+-----+-----+-----+-----+ 9060 CTGAAAAATGGCTCTCCGAAAACTACTGGCGACTACTATAGAACTGTTTCGTCGGTTCTGG T F T Q K A L M T A D D I L T K Q A R P - AAACAGTAAAAGGTTATTTTCCACATTACAGATGGTGTCCGACTATGTCAATATCCAAAT -----+-----+-----+-----+-----+-----+-----+-----+ 9120 TTTGTCTATTTTCCAAATAAAGGTGTAATGTCTACCACAAGGCTGATACAGTATAGTTTA N S K K V I F H I T D G V P T M S Y P I - TAATTTTAAATATACAGGAACGCGCAATCGTACAGAACTCAGCTGAATAATTTTAAAGC -----+-----+-----+-----+-----+-----+-----+-----+ 9180 ATTAAAAATTTATATGTCCTTGCTGCTTACATGCTCTTGAGTCGACTTATTAATAATTTGCG N F K Y T G T T Q S Y R T Q L N N F K A - AAAAACTCCAAATAGTAGCGGGATATTACTGGAGGACTTTGTTTACATGGTCAGCAGATGG -----+-----+-----+-----+-----+-----+-----+-----+ 9240 TTTTTGGAGTTTATCATGCGCCCTATAATGACCTCTGAAAACAATGTPACCAGTCGCTTACC K T P N S S G I L L E D F V T W S A D G -
---	------	---

PCT/US05/27239

273/487

Figure 101V

```
9241  TGAACATAAGATTGTTCTGTTGGAGATGGTGAAGTTATACAGATGTTTACGAAGAAACCTGT 9300
      +-----+-----+-----+-----+-----+-----+
      ACTTGTATTCTTAACAAGCACCTCTACCTTCAATAGTCTACAAATGCTTCTTTGGACA
      E H K I V R G D G E S Y Q M F T K K P V -
9301  AACAGACCAATACGGAGTTTCATCAAAATACCTTCAATCACCTCCATGGAGCAGAGAGCTAA 9360
      +-----+-----+-----+-----+-----+-----+
      TTGTCGTGTTATGCCCTCAAGTAGTTTATGAAGTTAGTGGAGGTACCTCGTCTCTCGATT
      T D Q Y G V H Q I L S I T S M E Q R A K -
9361  ATTAGTTTCAGCGGGATATAGTTCTATGGAACTGACTTGTATTTATTTATTTGGCGTGATAG 9420
      +-----+-----+-----+-----+-----+-----+
      TAATCAAAGTCGCCCTATATCCAAAGATACCTTGACTGAACATAAATAACCGCACATATC
      L V S A G Y R F Y G T D L Y L Y W R D S -
9421  TATTCAGCCTATCCATTAACTCTAGTACCGATGGATTGATTACCAACCATGGTGACCCCTAC 9480
      +-----+-----+-----+-----+-----+-----+
      ATAAGATCGGATAGGTAAATTGAGATCATGGCTAACCTAATGGTTGGTACCACCTGGGATG
      I L A Y P F N S S T D W I T N H G D P T -
9481  GACTTGGTATTATAACGGAAATATGGCTCAGGATGGCTATGATGTCTTCACCTGTTGGGGT 9540
      +-----+-----+-----+-----+-----+-----+
      CTGAACCATATAATTGCCCTTTATACGGAGTCCCTACCGATACACAGAGTGACAACCCCA
      T W Y Y N G N M A Q D G Y D V F T V G V -
9541  TGGTGTAAACGGGGATCCTGGTACGGATGAAGCAACGGCTACTAGATTATTCAGAGCAT 9600
      +-----+-----+-----+-----+-----+-----+
      ACCACATTTGCCCTTAGGACCATGCTACTTCTGTTGCCGATGATCTAAATACGTCTCGTA
      G V N G D P G T D E A T A T R F M Q S I -
9601  CTCTAGTTCTCCTGACAACTACACTAACGTAGAGATCCATCTCAGATTTTACAGAATT 9660
      +-----+-----+-----+-----+-----+-----+
      GAGATCAAGAGGACTGTTGATGTGATGTCATCGTCTAGGTAGAGTCTAAAATGTTCTTAA
```

Figure 101W

b S S S P D N Y T N V A D P S Q I L Q E L -  
9661 GAATCGCTACTTCTATATACTATCGTCAATGAGAAGAAATCTATCGAAATGGTACGATTAC 9720  
CTTAGCGATGAAGATATGATAGCAGTTACTCTCTCTTTAGATAGCTTTAGCATGCTAATG  
N R Y F Y T I V N E K K S I E N G T I T -  
9721 AGACCCGATGGGGAACCTAATTGATTTCCAAATTTGGGAGCAGATGGAGGTTTGTATCCAGC 9780  
TCTGGGCTACCCACTTGATTAACCTAAAGTTAACCCCTCGTCTACCTCCCAACCTAGGTCG  
D P M G E L I D F Q L G A D G R F D P A -  
9781 GGATACACTTTAACTGCAAAACGATGGTAGTTCGTTGGTCAATAATCTCCCTACTGGGGG 9840  
CCTAATGTGAATGACGTTTGTGTACCATCAAGCAACCACTTATTATAGGGGATACCCCC  
D Y T L T A N D G S S L V N N V P T G G -  
9841 ACCACAAAATGATGGTGGCTTGTCTAAATAAATGCAAAAGTCTTCTATGATACGACTGAGAA 9900  
TGGTGTCTTACTACCAACCGAACGATTTTCTACGTTTTCACAGATATGCTGACTCTT  
P Q N D G G L L K N A K V F Y D T T E K -  
9901 AAGATCGTGTACAGGTTTGTACCTTGGAAACGGGTGAAAGTTACATTGACTTATAA 9960  
TTCCTAAGCACATTTGTCCTCAACATGGAACCTTGCCCACTTTTTCATGTAACTGAATATT  
R I R V T G L Y L G T G E K V T L T Y N -  
9961 TGTTCGTTGAATGACCAATTTGTAAGCAATAAATCTATGACACGAAATGGTCCGAACAAC 10020  
ACAAGCGAACTTACTGGTTAAACATTCGTTATTAAAGTACTGTGCTTACCAGCTTGTG  
V R L N D Q F V S N K F Y D T N G R T T -  
10021 CCTACACCCCTAAGGAAGTAGAAAAGAACACAGTCCGGGACTTCCCGATTCCCTAAGATTCCG 10080

Figure 101X

GGATGTGGGATTCCTTCATCTTTCTGTGTCAACGGCTGAAGGGCTAAGGATTCCTAAGC  
L H P K E V E K N T V R D F P I P K I R -  
10081  
TGATGTACGAAAGTATCCAGAAATCACAATCCAAAAGAGAAAAACCTTGGTGAATTTGA  
ACTACATGCTTTCATAGGCTTTTAGTGTAAAGGTTTCTCTTTTGTGAACACACTTTAACT  
D V R K Y P E I T I P K E K K L G E I E -  
10141  
GTTTATTAGATCAATAAGAATGATAAAAAACCACCTGAGAGATGCGGCTTTAGTCTTCA  
CAAATAATCTAGTATTCTTACTATTTTTGGTACTCTCTACGCCAGAAATCAGAACT  
F I K I N K N D K K P L R D A V F S L Q -  
10201  
AAACAACATCCGGATTATCCAGATATTTATGGAGCTATTGATCAAAAATGGCACCTTATCA  
TTTTGTTAGGCCCTAATAGGCTATAAAATACCCTGATAACCTAGTTTACCGTGAATAGT  
K Q H P D Y P D I Y G A I D Q N G T Y Q -  
10261  
AAATGTGACAAACAGGTGAAGATGGTAAGTTGACCTTTAAATCTGTCTCAGATGGGAATA  
TTTACACTCTGTGCCACTTCTACCAATCAACTGGAAATTTTATAGACAGCTACCCCTTAT  
N V R T G E D G K L T F K N L S D G K Y -  
10321  
TCGATTATTGAAAATCTGAACCCAGCTGGTTATATAACCCGTTCAAAATAAGCCCTATCGT  
AGCTAATAAACCTTTTAAGACTTGTGACCAATATTTGGGCAAGTTTATTCGGATAGCA  
R L F E N S E P A G Y K P V Q N K P I V -  
10381  
TGCCCTCCAAAATAGTAATGGAGAGTCAGAGATGTGACTTAATCGTTCCACAAGATAT  
ACGGAAGGTTTATCATTTACCTCTTCAGTCTCTACACTGAAGTTAGCAAGGTGTTCTATA  
A F Q I V N G E V R D V T S I V P Q D I -



Figure 101Y

10441 ACCAGCGGTACGAGTTACGAATGATAAGCACTATATACAAATGAGCCAAATTCCTCC + 10500  
-----+-----+-----+-----+-----+-----+  
TGTCGCCCCAATGCTCAATGCTTACTATTCGTGATATAGTGTCTACTCGGTTAAGGAGG  
b P A G Y E F T N D K H Y I T N E P I P P -  
-----+-----+-----+-----+-----+-----+  
10501 AAAAAAGAAATATCCTCGAAGTGGTGGTATCGGAATGTGCCATTCTATCTCATAGGTTG + 10560  
-----+-----+-----+-----+-----+-----+  
TTTTTCTCTTATAGGAGCTTGACCAACCATAGCCCTTACAACGGTAAAGATAGACTATCCAAC  
K R E Y P R T G G I G M L P F Y L I G C -  
-----+-----+-----+-----+-----+-----+  
10561 CATGATGATGGAGGAGTTCTATATATATACACACGGAACATCCGTAAAGTGTAGCAATGAG + 10620  
-----+-----+-----+-----+-----+-----+  
GTACTACTACCCCTCCTCAAGATAATATGTGTGCCCTTTGTAGGCATTTCCATCGTTACTC  
M M M G G V L L Y T R K H P \*  
-----+-----+-----+-----+-----+-----+  
10621 AAATGATAATATCGATACCTCGAGCGGATATTTTAAAGTAGTACACTCAAGAGAGATTT + 10680  
-----+-----+-----+-----+-----+-----+  
TTTACTATTATAGCTATGAGACTCGCTATGAAATTTCTTCATCGTGAGTTCTCTCTTAAA  
AAGTTTACTTGTGAAACAGTTTCTTCGCCAAGTAAACACCATTGAAAGGGGAGATG + 10740  
-----+-----+-----+-----+-----+-----+  
TTCAAATGAACCACTTTTGTCAAAGAGCGGTTCAATTTGGTGGTAACTTTCCCTCTCTAC  
TTTTTCGAAAACCTGCACAGAAAAAGGATTATTTATTCATGTCATGTAATTCATTACATTGC + 10800  
-----+-----+-----+-----+-----+-----+  
10741 AAAAGCTTTTGAACGTGCTTTTTTCCATAATAATAACAGTACACATTAAAGTAATGTAACG  
-----+-----+-----+-----+-----+-----+  
TCACAGTTGATTTTAAGAGATATGAATAAGGAGAAATCATGAAATCAATCAACAAATTTT + 10860  
-----+-----+-----+-----+-----+-----+  
AGTGCAACTAAATTCCTCTATACTTATTCCTCTTTTACTACTTTTACTAGTTGTTTAAAA  
M K S I N K F L - orf4\_670, homologue of sp0463, LPXTG  
-----+-----+-----+-----+-----+-----+  
TAACAATGCTTGCTGCCCTTATTAATCTGACACGGAGTAGCCTGTTTTCAGCTGCAACAGTTT

[illegible]

PCT/US05/27239

278/487

Figure 101AA

11281 TTGGTCCTAATGGGCAAGTATTAAACAGTTCAAAAGCCGTACCTGCTCTTGTAACTCTTC + 11340  
AACCAGGATTACCCGTTTCATTAATTGTCCAAAGTTTTCGGCATGGACGAGAACATTGAGAAG  
C G P N G Q V L T G S K A V P A L V T L P -  
11341 CACTTGTTAACAATAATGGTACAGTAATTGATGCACATGTTTCCCTAAATAATTCATATA + 11400  
GTGAACAATTGTATTATACCATGTCTATTAACTACGTGACAAAAGGATTTTAAAGTATAT  
C L V N N N G T V I D A H V F P K N S Y N -  
11401 ATAAACCAGTTGTAGATAAAAAGAATTGCTGATACCTTTGAATTATAACGATCAAAAATGGTC + 11460  
TATTGGTCAACATCTATTTTCTTAACGACTATGAACCTTAATATGCTAGTTTACCAG  
C K P V V D K R I A D T L N Y N D Q N G L -  
11461 TGTCTATCGGTACTAAAATCCCATATGTTGTTAATAACAACAATTCCTCAAGTAATGCAACAT + 11520  
ACAGATAGCCATGATTTTGGGTATACACAAATTATGTTGTTAAGGTTCAATACGTTGTA  
C S I G T K I P Y V V N T T I P S N A T F -  
11521 TTGCAACTTCATTTTGGTCAGATGAAATGACAGAGGTCTAACTTATATAATGAAGATGTAA + 11580  
AACGTTGAAGTAAACCCAGTCTACTTTACTGTCTTCCAGATTGGAATATTACTTCTACATT  
C A T S F W S D E M T E G L T Y N E D V T -  
11581 CAATTACTTTGAATAATGTAGCTATGGATCAAGCTGATTAATGAAGTCACATAAGGAAATA + 11640  
GTTAATGAACCTTATTACATCGATACCTAGTTCGACTAATACTTCAGTGATTTCCCTTTAT  
C I T L N N V A M D Q A D Y E V T K G N N -  
11641 ATGGCTTTAACTTAAATAACAGACGAGGTTTAGCTAAATAATTAATGGAAGGATGCGAG + 11700  
TACCGAAATTGAATTTTAAATTGCTTCGTCCAAATCGATTTTAAATTACCATTCCTACGTC

Figure 101AB

[illegible]

280/487

Figure 101AC

```
CGTTCCCTTCAACCCCAATTTTGTGACCTTTCTATTATGGTGGCGAGGTAGTAGGAC
      K G K L G V K N W K D N N P A P I N P E -
12121 -----+-----+-----+-----+-----+-----+-----+
AAGAACACGCTGTAAACATACGGTAAAGTTTGTCAAGTAGACCAAAAGATACTC
      E P R V K T Y G K K F V K V D Q K D T R -
12180 -----+-----+-----+-----+-----+-----+-----+
TTCCTGGTGCACATTTTGTATGCCATTTTCAAAACAGTTTCATCTCGGTTTTTCTATGAG
      G T C T A G A A A T G C G C A G T T C G T T G T T A A A A A G C A G A T A C C A A T A T A T T C C C T T T A
12240 -----+-----+-----+-----+-----+-----+-----+
CAGATCTTTTACGGCTCAGCAACAATTTTTCGTCATCGTTATTTATATATAACGGAAT
      L E N A Q F V V K K A D S N K Y I A F K -
12241 -----+-----+-----+-----+-----+-----+-----+
AGTCACTGCACACAAGCTGCAGATGAAGAAAGCAGCAGCAACTGCAAAACAAAATGG
      S T A Q Q A A D E K A A A T A K Q K L D -
12300 -----+-----+-----+-----+-----+-----+-----+
TCAGTTGACGTGTGTTCGACGCTACTTTTTCGTCGTGCTGACGTTTGTGTTTAAAC
      A T G C A G C G G T A G C A G C T T A C A C A A T G C T G C A G A T A A G C C G C T C A A G C T A G T A G
12360 -----+-----+-----+-----+-----+-----+-----+
TACGTGCCCATCGTCGAATGTGTTACGACGCTCTATTTCGTCGGCGAGTTGAGATCATC
      A A V A A Y T N A A D K Q A A Q A L V D -
12361 -----+-----+-----+-----+-----+-----+-----+
ATCAAGCACAGCAGCAATACAAATAGTAGCTTACAAAGAACCAAAATTTGGTTATGTTGAAG
      Q A Q Q E Y N V A Y K E A K F G Y V E V -
12420 -----+-----+-----+-----+-----+-----+-----+
TAGTTGTTGCTGTTCTTATGTTACATCGAATGTTTCTTCGGTTTAAACCAATACAACTTC
      T A G C T G G A A A A G A G C A A T G G T T C T T A C T T C T A A T A G G A T G G T C A A T T C C A A A T T T
12480 -----+-----+-----+-----+-----+-----+-----+
ATCGACCTTTTCTACTTCGTTACCAAGAAATGAGATTATGCCTACGAGTTAAGGTTTAAA
      A G K D E A M V L T S N T D G Q F Q I S -
```

Figure 101AD

12481 CAGGTCCTTGGCTGGTACTTTATATAATTAGAAGAAATTAAAGCTCCAGAAAGGTTTTCGGA 12540  
-----+-----+-----+-----+-----+-----+-----+-----+  
GTCCAGAACCGACGACCAATGAATATTTAAATCTTCTCTTTTAAATTTTCGAGGGTCTTCCAAAAACGCT  
G L A A G T Y K L E E I K A P E G F A K -  
AAATTGATGATGATAGAAATTTGTTGTGGAGCAGGTTCTTGGAAATCAAGGTGAGTTTAATT 12600  
-----+-----+-----+-----+-----+-----+-----+-----+  
TTTAACTACTACATCTTAAACAACACCTCGTCCAAAGAACCTTAGTTCCTACTCAAAATTA  
I D D V E F V V G A G S W N Q G E F N Y -  
ACTTAAAGATGTTCAAAGAATGACGCTACAAAAGTAGTCAACAACAAAAATCCTATCC 12660  
-----+-----+-----+-----+-----+-----+-----+-----+  
TGAATTTTCTACAAGTTTCTTACTACGGGATGTTTTCATCAGTTGTTTTTTAGTCATAGG  
L K D V Q K N D A T K V V N K K I T I P -  
CAGAAACGGGTGGTATTGGTACAAATATCTTTTGTGTAGCGGGGGCTCGGATTATGGGTA 12720  
-----+-----+-----+-----+-----+-----+-----+-----+  
GTGTTTGGCCACCATAACCATGTTAATAGAAACGACATCGCCCCCGAGCTAATACCCAT  
Q T G G I G T I I F A V A G A A I M G I -  
TTGCAGTGTACGCATATGTTAAACAAACAAGATGAGGATCAACTTGCTTAAGTAAGAG 12780  
-----+-----+-----+-----+-----+-----+-----+-----+  
AAGCTCACATCGGTATACAATTTTGTGTTTCTTACTCTCCTAGTTGAACGAATTCATCTC  
A V Y A Y V K N N K D E D Q L A \*  
AGAAAGGAGCCATTTGATGACAAATGCAGAAAAATGCAGAAAAATGATTAGTCGTATCTCTTT 12840  
-----+-----+-----+-----+-----+-----+-----+-----+  
TCTTTCTCGGTAACTACTGTTACGTCCTTTTACGTCCTTTTACTAATATCAGCATAGAAGAAA  
M T M Q K M Q K M I S R I F F -orf5\_670, homologue of sp0464, LPXTG  
GTTATGGCTCTGTGTTTTTCTCTGTATGGGTGCACATGCGAGTCCAAAGCGCAAGAGAT 12900  
-----+-----+-----+-----+-----+-----+-----+-----+  
CAATACCGAGACACAAAAAGAGAAACATACCCCGTGTACCTCAGGTTCCGGTCTCTCTTA

PCT/US05/27239 82/487

Figure 101AE

a V M A L C F S L V W G A H A V Q A Q E D -  
12901 CACACGTTGGTCTTGGCAATTGGGAGAACTATCAGGAGGTGGTTAGTCAATTGCCATCTCGT  
GTGCGCAACGAGAACGTTAACTCTTGTAGTCTCTCCACCAATCAGTTAACGGTAGAGCA  
12960  
a H T L V L Q L E N Y Q E V V S Q L P S R -  
12961 GATGGTCATCGGTTGCAAGTATGGAAGTTGGATGATTCGTATTCCTATGATGATCGGGTG  
CTACAGTAGCCACGTTTCATACCTTCAACCTACTAAGCATAAGGATACTACTAGCCCCAC  
13020  
a D G H R L Q V W K L D D S Y S Y D D R V -  
13021 CAAATTGTAAGAGACTTGCCATTGCGGATGAGAAATAAATTTCTTCTTCAAAAAGACT  
GTTTAACTCTCTGTAACGTAAGCACCCCTACTCTTATTTGAAAGAGAAAGTTTTCGTA  
13080  
a Q I V R D L H S W D E N K L S S F K K T -  
13081 TCGTTGAGATGACCTTCCTTGAGATATCAGATGGAATGATCTCATATCCAAATGGTCTT  
AGCAACTCTACTGGAAGGAACCTCTTAGTCTAATTCATAGAGATTAAGTTTACCAGAA  
13140  
a S F E M T F L E N Q I E V S H I P N G L -  
13141 TACTATGTTGGTCTATTATCCAGACGGATGCGGTTTCTTATCCAGCTGAATTCCTTTT  
ATGATACAAGCGAGATAATAGGTCTGCTACGCCCAAGAAATAGGTGACTTAAAGAAAAA  
13200  
a Y Y V R S I I Q T D A V S Y P A E F L F -  
13201 GAAATGACAGATCAACGGTAGAGCCTTTGGTCAATGTAGCGAAAAAACAGATACAATG  
CTTACTGTCTAGTTTGCCATCTCGGAACCCAGTAACATCGCTTTTGTCTATGTTAC  
13260  
a E M T D Q T V E P L V I V A K K T D T M -  
13261 ACAACAAAGGTGAAGCTGATAAAGGTGGATCAAGACACCAATCGGTTGGAGGTGTCGGC  
13320

PCT/US05/27239 283/487

Figure 101AF

```

a      T G T G T T T C C A C T T C G A C T A T T T C C A C C T A G T T C G T G T T A G C G A A C C T C C C A C A C C G
      T T K V K L I K V D Q D H N R L E G V G -
      T T T A A T T G G T A T C A G T A G C A A G A G A T G G T T C T G A A A A A G A G G T T C C C T G A T T G G A G A A
13321  +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      A A T T T A C C A T A G T C A T C G T T C T C T A C C A A G A G T T T T T C C A A G G G A A C T A A C C T C T T
a      F K L V S V A R D G S E K E V P L I G E -
      T A C C G T T A C A G T T C T T C T G G T C A A G T A G G G A G A C T C T C T A T A C T A T G A T A A A A A T G G A G A G
13381  +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      A T G G C A A T G T C A A G A G A C C A G T T C A T C C C T C T T G A G A G A T A T G A C T A T T T T T A C C T C T C
a      Y R Y S S S G Q V G R T L Y T D K N G E -
      A T T T T T G T G A C A A A T C T C T C T T G G G A A C T A T C G T T T C A A G A G G T G G A G C C A C T G G C A
13441  +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      T A A A A C A C T G T T T A G A A G G A G A A C C C T T G A T A G A C A A A G T T C C T C C A C C T C G G T G A C C G T
a      I F V T N L P L G N Y R F K E V E P L A -
      G G T A T G T G T T A C G A C G C T G G A T A C G G A T G T C C A G C T G G T A G A T C A T C A G C T G G T G A C G
13501  +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      C C G A T A C G A C A A T G C T G C G A C C T A T G C C T A C A G G T C G A C C A T C T A G T A G T C G A C C A C T G C
a      G Y A V T T L D T D V Q L V D H Q L V T -
      A T T A C G G T T G T C A A T C A G A A A T T A C C A C G T G G C A A T G T T A C T T T A T G A A G G T G G A T G G T
13561  +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      T A A T G C C A A C A G T T A G T C T T T A T G G T G C A C C G T T A C A C T G A A A T A C T T C C A C C T A C C A
a      I T V V N Q K L P R G N V D F M K V D G -
      C G G A C C A A T A C C T C T C T T C A A G G G C A A T G T T C A A A G T C A T G A A A G A A A A G C G G A C A C
13621  +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      G C C T G G T A T G G A G A A G T T C C C G T T A C A A G T T T C A G T A C T T T C T T C T T T C G C C T G T G
a      R T N T S L Q G A M F K V M K E E S G H -

```



Figure 101AG

13681	a	TATACTCCTGTTCTTCAAAATAGGTAAAGGAGTAGTGTGTTAAACATCAGGGAAAGATGGTCTGCT	13740
		ATATGAGGACAAAGAAGTTTACCATTCTCTTCATCAACATTTGTAGTCCCTTTCTTCTACCAAGCA	
	a	Y T P V L Q N G K E V V V T S G K D G R -	
13741		TTCCGAGTGGAAAGGTCTAGAGTATGGGACATACATAATTTATGGGAGCTCCAAGCTCCAAGT	13800
		AAGGCTCACCTTCAGATCTCATACCTGTATGATAAATAACCTCGAGGTTGCGAGGTTGA	
	a	F R V E G L E Y G T Y Y L W E L Q A P T -	
13801		GGTTATGTTCAANTTAACATCGCTGTTCTCTTTACATCGGGAAAGATACCTCGTAAGGAA	13860
		CCAAATACAAGTTAAATGTAGCGGACAAAGGAAATGTTAGCCCTTTCTATGAGCATTCCTT	
	a	G Y V Q L T S P V S F T I G K D T R K E -	
13861		CTGTTAACAGTGGTTAAATAACAAGCGACCAAGGATGTGTCAGATACAGGGGAA	13920
		GACCATGTGCACCAATTTTATTTATGTTCTGGTGGTGCCTAACTACAGGTCTATGTCCCTT	
	a	L V T V V K N N K R P R I D V P D T G E -	
13921		GAAACCTTGTAATATCTTGATGCTTGTGTCATTTTGTGTTTGGAGTGGTATTATATCTT	13980
		CTTTGGAACATATAGAACTACCAACAACCGTAAACACAAACCAATCACCATAATAGAA	
	a	E T L Y I L M L V A I L L F G S G Y Y L -	
13981		ACGAAAAAACCAATAACTGATATTCATGTACATCATATTATGAAAAAGATAGCAGGCTGA	14040
		TGCCTTTTGTGGTTTATTGACTATAAGTTACATGTAGTAATACCTTTTCTATCGTCCGACT	
	a	T K K P N N *	
14041		AGGGAAGACCAGAGTACTCTCAGGTGATGTTTAATCAGGAATCATGGTGATGGGCATGAA	14100
		TCCCTTCTGGTCTCATGAGACTCCACTCAATTAGTCTCTTAGTACCCTACACCGTACTTT	

Figure 101AH

14101 TCACAATAACGGGATATGAGCGTGGCAGATGTGCCAGCCCTCATGTGGGTATTGTTTG + 14160  
-----+-----+-----+-----+-----+  
AGTGTATTGGCTATACCTCGGACCGCTCTAACACGGTGGAGTAACACCCCAATAACAAAC  
14161 TAAACGATAGGACTGGTCTGGTAATCATTTTAGGAATGGACAGGACTGGGATTCGTGAT + 14220  
-----+-----+-----+-----+-----+  
ATTTGCTATCCTGACCCAGACCATTAAGTAAATCCTTACCTGCTCCTGACCCCTAAGACTAA  
14221 TAAATGGATGGTGAATCAGAAAGAAATGAGATTTTCTCGTTTCTCTTAGCAGATAGGAT + 14280  
-----+-----+-----+-----+-----+  
ATTTACCTACCACTTAGTCTTCTTTACTCTCTAAAGAGCAAGAGAAATCCTCTATCCTA  
14281 TGTCTGTAGGAAAAGCGATATAAATGATGAGTTTGAAGATAAAGGGATGCTGATATAAAAT + 14340  
-----+-----+-----+-----+-----+  
ACAGACAAATCCTTTTTCGCTATTTTACTACTCAAACTCTATTTTCCCTACGACTATTTTAA  
M L I K M -orf6\_670, homologue of sp0466, sortase  
GGTAAAAACAAAAAGCAAAAACGAATAATCTCCTATTAGGAGTGGTATTTTCATTGG + 14400  
-----+-----+-----+-----+-----+  
CCATTTTGTGTTTTTTCGTTTTTTCCTTTATTAGAGGATAATCCTCACCATAAAAAGTAACC  
V K T K K Q K R N N L L L G V V F F I G -  
14401 AATGGCGGTAATGGCGTATCCGCTGGTGTCTCGTCTTATTATCGAGTGGATCAAAATCA + 14460  
-----+-----+-----+-----+-----+  
TTACCGCCATACCGCATAGCGGACCCACAGAGCGGAACATAATAGCTCACCCTAGTTTACT  
M A V M A Y P L V S R L Y Y R V E S N Q -  
14461 ACAAAATTGCTGACTTTGATAAGGAAAACCAACGTTGGATGAGGCTGACATGATGAACG + 14520  
-----+-----+-----+-----+-----+  
TGTTTAACGACTGAACCTATTCCTTTTTCGTTGCAACCTACTCCGACTGTAACTACTTGC  
Q I A D F D K E K A T L D E A D I D E R -  
AATGAAATGGCACAAGCCTTCAATGACTCTTTTGAATAAATGATGAGTGGGATCCTTG

14521	TTACTTTTAA	CCGCTGTT	CGGAAGTT	TACTGAGAAAC	CTTATTACAT	CATCACT	CACCGCTAGGAAC	14580
b	M K L A Q A F N D S L N N V V S G D P W -							
14581	GTCCGAAGAA	ATGAGAAAA	AGGGCGAG	CAGAGTAG	TGCACGTAT	TGCAGTAAT	CCATGA	14640
	CAGCCTTC	TTTACTTC	TTTTTTCC	CGCTCTCT	CATACGTG	CATACAA	CTTTAGGTACT	
b	S E E M K K K G R A E Y A R M L E I H E -							
14641	GGCGATGGG	CATGTGAA	ATCCCGTT	ATTGACG	TGTGATTT	GCGGTTT	ATGCTGGTAC	14700
	CGCCTAC	CCCCGTAC	ACCTTT	AGGGGCA	TAACTG	CACCTAA	ACGCCAAATACGACCATG	
b	R M G H V E I P V I D V D L P V Y A G T -							
14701	TGCTGAAG	AGGTATT	GCAGCA	AGGGGCT	GGGCAT	CTAGAGG	GAACCTTCTCTGCGGATCGG	14760
	ACGACTT	CTCCATA	ACGCTG	TCTCCCCG	ACCCGTAG	TCTCCCTT	GAAAGACGCGCTAGCC	
b	A E E V L Q Q G A G H L E G T S L P I G -							
14761	AGGCAAT	TCGACCC	ATCGGGTG	ATTACGG	CACATAC	AGTTTGC	CAACACGCTAAGATGTT	14820
	TCCGTTA	AGCTGG	TACGCC	ACTAAT	GCCGTG	TATGTCC	AAACGGTGTGCTATCTTACAA	
b	G N S T H A V I T A H T G L P T A K M F -							
14821	TACGGAT	TTGACCA	AACTTAA	AGTTGGG	ATAAGTTT	TATGTGC	ACATATATCAAGGAAGT	14880
	ATGCCAA	AACTGG	TTGAATT	CAACCC	CTATTCAA	AATAC	ACGGTGTATAGTTCCTTCA	
b	T D L T K L K V G D K F Y V H N I K E V -							
14881	GATGGCT	ATCACT	GGATCA	AGTAA	AGGTG	ATTGAC	CGGACGAACCTTTGATGATTTATT	14940
	CTACCG	GATAGT	TACCTAG	TTCATTT	CCACTA	CTCGCT	GCTTGAAACACTACTAATAA	
b	M A Y O V D O V K V I E P T N F D D L L L -							

287/487

Figure 101AJ

14941 GATTGTACAGGTCATGATTATGTGACCTTGTGACTTGTAGCCCATACATGATCAATAC 15000  
CTAACATGGTCCAGTACTAATACACTGGAACGACTGAACATGCGGTATGTACTAGTTATG  
b I V P G H D Y V T L L T C T P Y M I N T -  
15001 CCATCGTCTATTGTTTCGGGGCATCGGATACCGTACGTACGAGAGGTGAGGAAGAATT 15060  
GGTAGCAGATAACCAAGCCCCCGTAGCCTATGGCATGCACTGCTCCCAACTCCTTCTTAA  
b H R L L V R G H R I P Y V A E V E E F -  
15061 TATTGCAGCAAACTCAGTCACTCTATCGCTACCTGTTTATGTGGCAGTTGGTTT 15120  
ATAACGTCGTTGTTGAGTCAGTAGAGATAGCGATGGACAAAATACACCGTCAACCAAA  
b I A A N K L S H L Y R Y L F Y V A V G L -  
15121 GATTGTGATTTCTTTATGGATTATTCGACGCTTTCGCGAAGAGAAAACCAACCGGAAAA 15180  
CTAACACTAAGAAAATACCTAATAAGCTCGAAGCGGTTCTCTTTTGTGGCTTTT  
b I V I L L W I I R R L R K K K Q P E K -  
15181 GGCTTTGAAGGGCTGAAAGCAGCAAGGAAGGAAGTGAAGTGGAGGATGGACAACAGTA 15240  
CCGAAACTCCGGGACTTTCGTGCTCCTTCTCTCCTTCCACTTCCACCTCCCTGTTGTCAT  
b A L K A L K A A R K E V K V E D G Q Q \* -  
15241 GACGTTACGAAAAAAGGCACAAAAAGAAAGAACATCCGCTGATCCTTCTTCTGATTT 15300  
CTGCAAGTCTTTTTCGGTGTGTTTCTTCTTCTTGTAGGCGACTAGGAAGAAGACTAAA  
TCTTAGTAGGATTCGGCGTTGGGATATATCCATTGGTCTCGTTATTATTATTCGTATTG  
15301 AGAATCATCTAAGCGCAAGCGCTATATAGGTAACCAAGACAGCAATATAATAGCATAAC 15360  
V S R Y Y R I E -orf7\_670, homologue of sp0467, sortase

PCT/US05/27239

288/487

Figure 101AK

```
15361 AGTCAAACGAGGTTATTAAAGAGTTTGATGAGACGGTTTCCAGATGGATAAGGCAGAAC 15420
-----+-----+-----+-----+-----+-----+-----+-----+
TCAGTTTGCTCCCAATAATTTCCTCAAACTACTCTGCGCAAGGGTCTACCTATTCCGTCCTTG
c S N E V I K E F D E T V S Q M D K A E L -

15421 TTGAGGAGCGTTGGCGCTTGCTCAAGCCTCAATGCGACCTTGAACCATCTGAAATTC 15480
-----+-----+-----+-----+-----+-----+-----+-----+
AACTCCTCGCAACCGCGAACCGAGTTCGGAAAGTTACGCTGGAACCTTGGTAGACTTTTAAG
c E E R W R L A Q A F N A T L K P S E I L -

15481 TTGATCCTTTTACAGAGCAAGAGAAAAAGAAAGCGCTCTCAGAAATATGCCAATATGCTAA 15540
-----+-----+-----+-----+-----+-----+-----+-----+
AACTAGGAAAAATGTCTCGTTCGTTCTTTTCTTTTCCGCGAGAGTCTTATACGGTTATACGATT
c D P F T E Q E K K K G V S E Y A N M L K -

15541 AGTCCATGAGCGGATTGGCTATGTGGAATTCCTCGGATTGATCAGGAAATCCGATGT 15600
-----+-----+-----+-----+-----+-----+-----+-----+
TCCAGGTACTCGCCTAACCGATACACCTTTAAGGACGCTAACTAGTCTCTTAAAGGCTACA
c V H E R I G Y V E I P A I D Q E I P M Y -

15601 ATGTCCGAACGAGTGAGGAAATTCTTCAGAAAGGCGCGAGGATTGCTAGAGGGAGCTTCGT 15660
-----+-----+-----+-----+-----+-----+-----+-----+
TACAGCCTTGCTCAGTCTCTTTAAGAAGTCTCCCGGCTTAACGATCTCCCTCGAAGCA
c V G T S E E I L Q K G A G L L E G A S L -

15661 TACCGGTTGGTGGTGAATAATACCCACACAGTTGTCTACTGCTCATAGAGGATTACCGACGG 15720
-----+-----+-----+-----+-----+-----+-----+-----+
ATGGCCAACCAACCACTTTTATGGGTGTGTCAACAGTGACGAGTATCTCCTAATGGCTGCC
c P V G G E N T H T V V T A H R G L P T A -

15721 CAGAACTGTTTAGTCAATTGGATAAGATGAAAAAGGGGATGCTTTTATCTTCACGTTT 15780
-----+-----+-----+-----+-----+-----+-----+-----+
GTCCTTGACAAATCAGTTAACCTATTCTACTTTTTTCCCTTACAGAAAAATAGAAAGTGCAAA
```

PCT/US05/27239

289/487

Figure 101A

c E L F S Q L D K K K G D V F Y L H V L -  
15781 TAGACAGGTGTTGGCCTACCAAGTGGATCAGATTTTGACGGTTGAGCCAAATGACTTTG + 15840  
ATCTGGTCACAACCGGATGGTTCACCTAGTCTAAAACCTGCCAACTCGGTTTACTGAAAC  
c D Q V L A Y Q V D Q I L T V E P N D F E -  
15841 AGCCTGTCTTGATTCACATGGGGAGATTATGCGACCTTGTGTGACCTGTACACCGTATA + 15900  
TCGGACAGAACTAAGTTGTACCCCTCTAATAGCTGGACAACTGGACATGTGGCATAT  
c P V L I Q H G E D Y A T L L T C T P Y M -  
15901 TGATTACAGTCATCGTCTGTGGTACGTTGGGAGCGGATTCCGTATACGGCACCAATTG + 15960  
ACTAATTGTCAGTAGCAGACAAACCATGCACCTTCCCTAGGCATATGCCGTGGTTAAC  
c I N S H R L L V R G K R I P Y T A P I A -  
15961 CAGAGCGAATCGAGCGGTGAGAGCGGTGGGAAATTCGTGTGTGTTATTGCTACCGG + 16020  
GTCTCGCTTAGCTCGGCACCTCTCTGACCCGTTAAGACCAACCAATAACGATCGCC  
c E R N R A V R E R G Q F W L L L L A A -  
16021 CGTTGGTTATGATTCCTGTTATGAGTTACGGGGGTGTATCGTCATCGTCGCATTGTCAAAG + 16080  
GCAACCAATACTAAGACATAACTCAATGCCCCACATAGCAGTAGCAGCGTAACAGTTTC  
c L V M I L V L S Y G V Y R H R R I V K G -  
16081 GGCTAGAAAACAAATTGGAGGAGCATCATGTCAAGGCTAAGCTACAGAAATTACTAGGG + 16140  
CCGATCTTTTGTAACTCCTCGTAGTACAGTTCCGATTCGATGTCCTTAATGATCCC  
a M S K A K L Q K L L G -orf8\_670, homologue of sp0468, sortase  
c L E K Q L E E H H V K G \*

290/487

Figure 101AM

16141 TATTGCTGATGCTGGTACCATTTGGTATTCCTGTTTATTGTTTGGGAGATGGTGTTA 16200  
-----+-----+-----+-----+-----+-----+  
16142 ATAAACGACTAGACCATCGTAACCACTAAGGACAAATAACAAAACCGTCTTACCAAT  
-----+-----+-----+-----+-----+-----+  
a Y L L M L V A L V I P V Y C F G Q M V L -  
-----+-----+-----+-----+-----+-----+  
16201 CAGTCTTTAGGACAAAGTAAAGTCAATGAGATATTTTCAATCTGTGACGGCCGACAGT 16260  
-----+-----+-----+-----+-----+-----+  
GTCAGAAATCCTGTTTCATTTTCCAGTACTCTATAAAAGTCTTAGACACATGCCCGGTGTCA  
-----+-----+-----+-----+-----+-----+  
a Q S L G Q V K G H E I F S E S V T A D S -  
-----+-----+-----+-----+-----+-----+  
16261 TACCAAGACCAATTGCAACGGTGGTGTGATTACAAATCAACGCTGGATTGCGCAAAATCGT 16320  
-----+-----+-----+-----+-----+-----+  
ATGGTTCCTGTTAACGTTGCCAGCGAACTAATGTTAGTTGGGAACCTAAGCGTTTGTAGCA  
-----+-----+-----+-----+-----+-----+  
a Y Q E Q L Q R S L D Y N Q R L D S Q N R -  
-----+-----+-----+-----+-----+-----+  
16321 ATTGTAGATCCCTTTTGGCGGAAGGTATGAGTAAATACCAAGTCTGTGACGATCCT 16380  
-----+-----+-----+-----+-----+-----+  
TAACATCTAGGAAAAAACCGCTTCCCATCTCCATTTAATGGTTCACAGACTGCTAGGA  
-----+-----+-----+-----+-----+-----+  
a I V D P F L A E G Y E V N Y Q V S D D P -  
-----+-----+-----+-----+-----+-----+  
16381 GATGCAGTCTACGGCTATTGTCGATTCGAGTTGGAAATCATGAGCCAGTTTATCTA 16440  
-----+-----+-----+-----+-----+-----+  
CTACGTCAGATGCCGATAACAGCTAAGGCTCAACCTTTAGTACCTCGGTCAAATAGAT  
-----+-----+-----+-----+-----+-----+  
a D A V Y G Y L S I P S L E I M E P V Y L -  
-----+-----+-----+-----+-----+-----+  
16441 GGAGCGGATTACCATCATTTAGCAATGGGTTGGCCCATGTGGATGGACGCCCTTCCT 16500  
-----+-----+-----+-----+-----+-----+  
CCTCGCTAATGTTAGTAATTCGTTACCCCAACCGGGTACACCTACCTGCGGAGAAAGGA  
-----+-----+-----+-----+-----+-----+  
a G A D Y H H L A M G L A H V D G T P L P -  
-----+-----+-----+-----+-----+-----+  
16501 GTTGAGGGAAGGATTCGTTCACTGATTTGTTGGGACCGTGCAGAACCAAGCCATGTC 16560  
-----+-----+-----+-----+-----+-----+  
CAACTCCCTTTTCCCTAAGCAAGTCACTAACGACCCGTTGGACGTTCTTGGTTCCGTACAG  
-----+-----+-----+-----+-----+-----+

Figure 101AM

a V E G K G I R S V I A G H R A E P S H V -  
TTTTTCGGCCATTGGATCAGCTAAAGTTGGAGATGCTCTTTATTATGATAATGGCCAG 16561  
-----+-----+-----+-----+-----+  
AAAAAGCGGTAAACCTAGTCGATTTTCAACCTCTACGAGAAATAATACTATTACCGGTC  
a F F R H L D Q L K V G D A L Y Y D N G Q -  
GAAATTGTAGAAATATCAGATGATGGACACAGAGATTATTTACCGTCGGAATGGGAAAAA 16621  
-----+-----+-----+-----+-----+  
CTTTAACATCTTATAGTCTACTACCTGTGTCTCTAATAAAATGGCAGCCTTACCCCTTTT  
a E I V E Y Q M M D T E I I L P S E W E K -  
TTAGAAATCGGTTAGCTCTAAAAATATCATGACCTTGATAACCTGGATCCGATTCCTTACC 16681  
-----+-----+-----+-----+-----+  
AATCTAGCCAAATCGAGATTTTATATAGTACTGGAACHTATTGGACGCTAGGCTAAGGATGG  
a L E S V S S K N I M T L I T C D P I P T -  
TTTAATAAACGCTTATTAGTGAATTTTGAACGAGTCGCTGTTTATCAAAAAATCAGATCCA 16741  
-----+-----+-----+-----+-----+  
AAATTATTGGGAATAATCACTTAAACCTTGCTCAGCGACAAATAGTTTATTAGTCTAGGT  
a F N K R L L V N F E R V A V Y Q K S D P -  
CAACAGCTGCAGTTCGAGGGTTGCTTTACGAAAGAGGACAACTGTATCGCGTGT 16801  
-----+-----+-----+-----+-----+  
GTTTGTGACGTCACGCTCCCAACGAAATGCTTTCTCTCTGTAGACATAGCGCACAA  
a Q T A A V A R V A F T K E G Q S V S R V -  
GCAACTCTCAATGGTTGTACCGTGGGCTAGTGGTACTGGCATTTCTGGGAATCCTGTTT 16861  
-----+-----+-----+-----+-----+  
CGTTGGAGAGTTACCAACATGGCAGCCGATCACCATGACCGTAAAGACCCCTTAGGACAAA  
a A T S Q W L Y R G L V V L A F L G I L F -  
GTTTGTGGAAGCTAGCACGTTTACTACGAGGGAATAAAAAAGAAATGAAGGAAAGCTA 16921  
-----+-----+-----+-----+-----+  
-----+-----+-----+-----+-----+ 16980



PCT/US05/27239 292/487

Figure 101AN

a . V L W K L A R L L R G K \*  
CAAACACCTTCGATCGTGCAGAAATGATGCTCCCTTTATTTTCTTTACTTTCCCTTTTCGAT  
16981 AGGCTGTTTCCTTTTTCGGGCTCTTTTGTCAACTGTAGTGGGTGAAAAAAGCTAAGCTCG  
TCCGACAAGGAAAAAGGCGGAGAAACAGTTGACATCACCCCACTTTTTCGATTTCGAGC 17040  
17041 AGAAAGGACAAAATTTTGTCCCTTCTTTTGTGATATTCAGAGCGATAAAAATCCGTTTTTT  
TCTTTCCTGTTTAAACAGGAAGAAAAACATATAAGTCTCGCTATTTTTCGCAAAAAA 17100  
GAAGTTTTCAAA  
17101 -----+--- 17112  
CTTCAAAAGTTT

293/487

M1, strain 2580

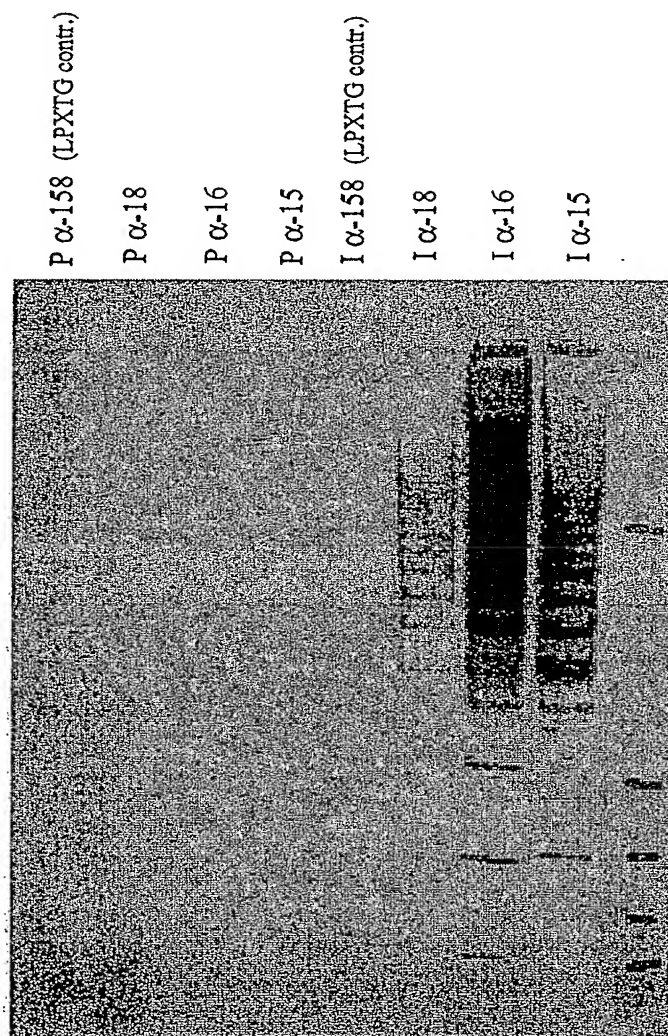


Figure 102

**LEGEND:**I  $\alpha$ -#: immune serum anti-#P  $\alpha$ -#: pre-immune serum anti-#

PCT/US2005/027239 294/487

M1, strain 2913

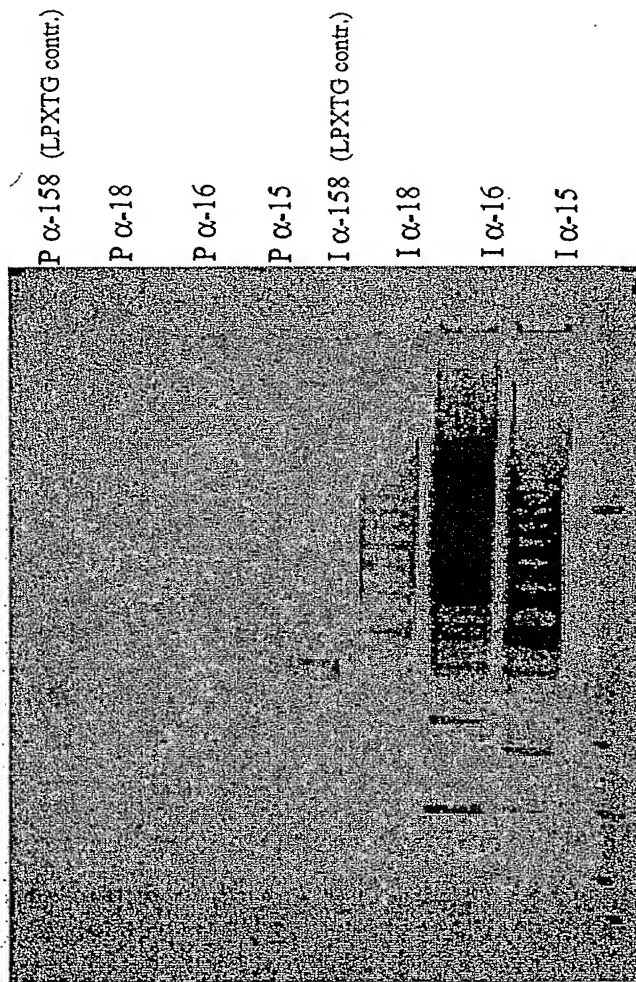


Figure 103

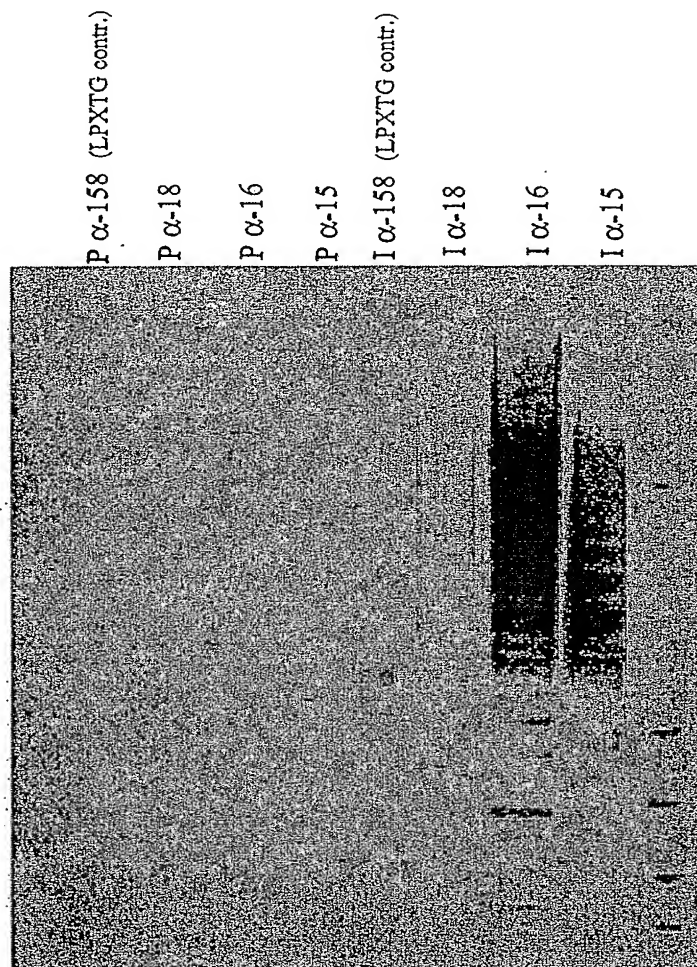
**LEGEND:**

I α-#: immune serum anti-#

P α-#: pre-immune serum anti-#

295/487

M1, strain 3280

**LEGEND:**

I α-#: immune serum anti-#

P α-#: pre-immune serum anti-#

Figure 104

296/487

## M1 strain 3348

P  $\alpha$ -158 (LPXTG contr.)  
P  $\alpha$ -18  
P  $\alpha$ -16  
P  $\alpha$ -15  
I  $\alpha$ -158 (LPXTG contr.)  
I  $\alpha$ -18  
I  $\alpha$ -16  
I  $\alpha$ -15

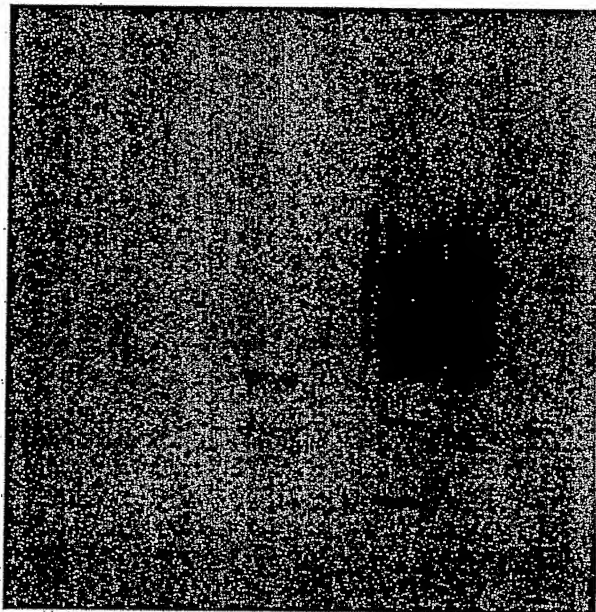


Figure 105

## M1 strain 2719

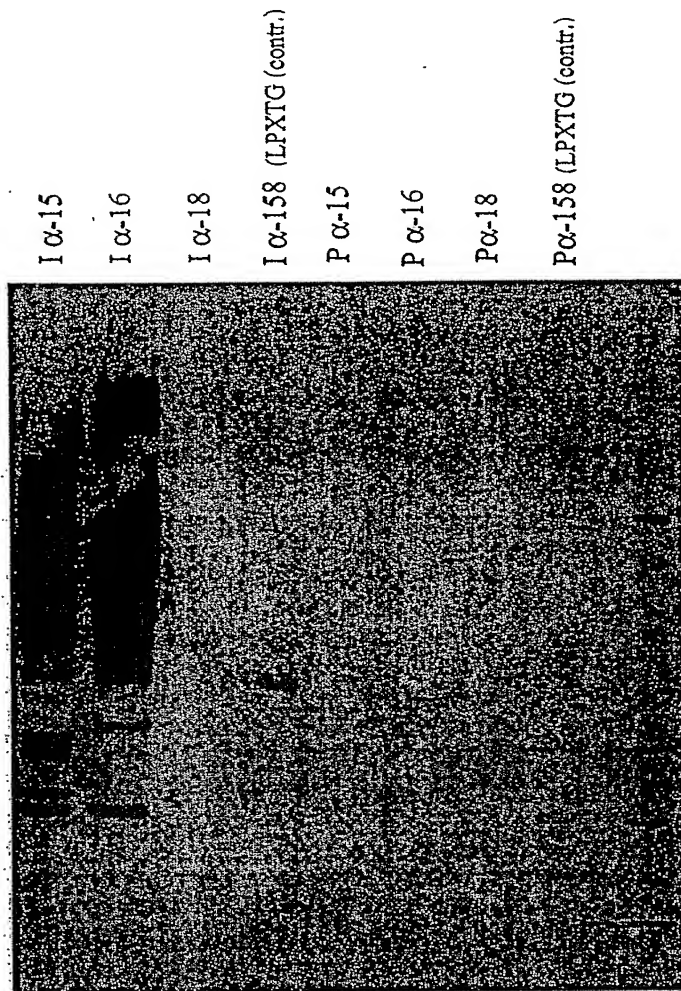


Figure 106

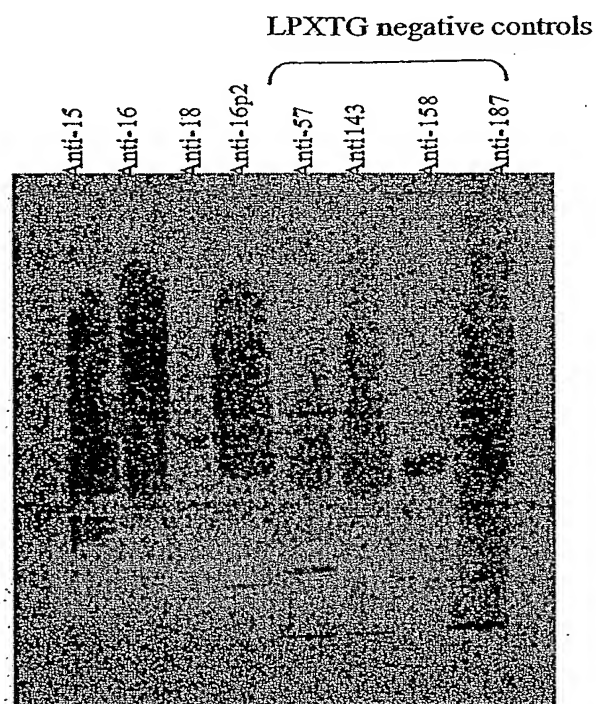
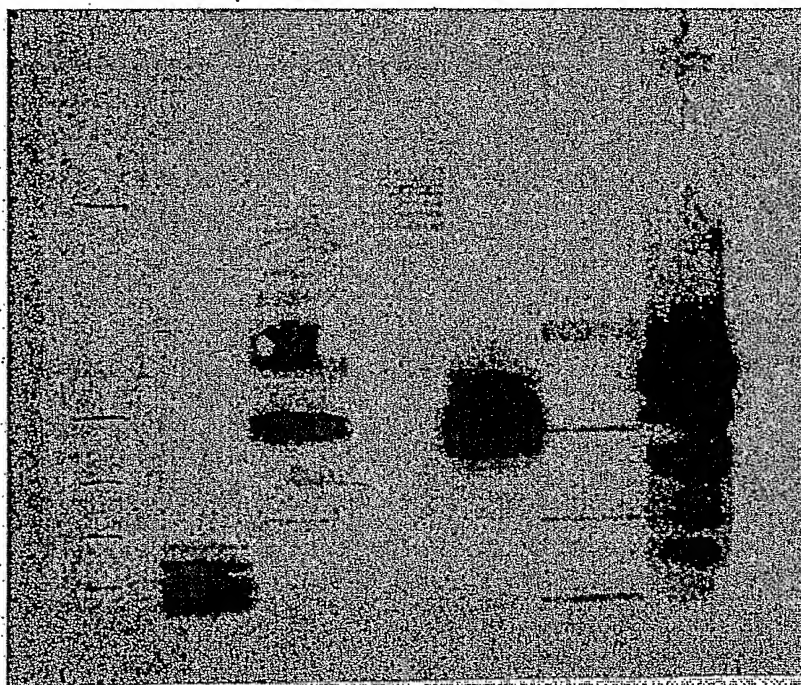
**Western blot Western blot on fraction enriched in surface proteins of M1 (SF370)**

Figure 107

## Western blot on fraction enriched in surface proteins of M12 (2728)

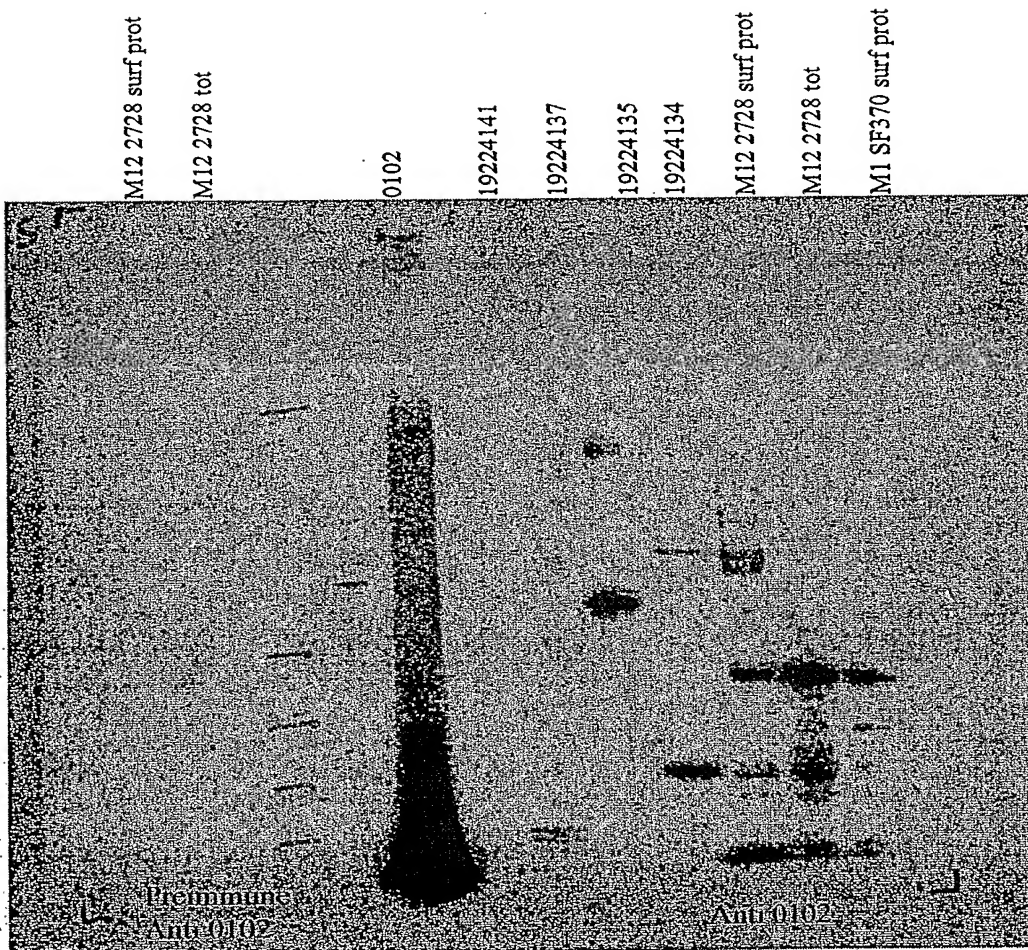
Anti-19224134  
Anti-19224135  
Anti-19224137  
Anti-19224141  
Anti-0102  
Anti-158 (control)



**Note:** 0102 corresponds to SpyM3\_0102 in MGAS315 (M3). This protein is pretty much identical to the M12 19224139 -see also gel "WB M12 (23-11-4b2a)"-, the M18 spyM18\_0130 and the M5 orf 82. Here the antiserum raised against 0102 recognizes high-weight molecular structures in M12 (strain isolate 2728). This means that in M12 three out of the five LPKTG proteins form high-molecular weight structures:

Figure 108

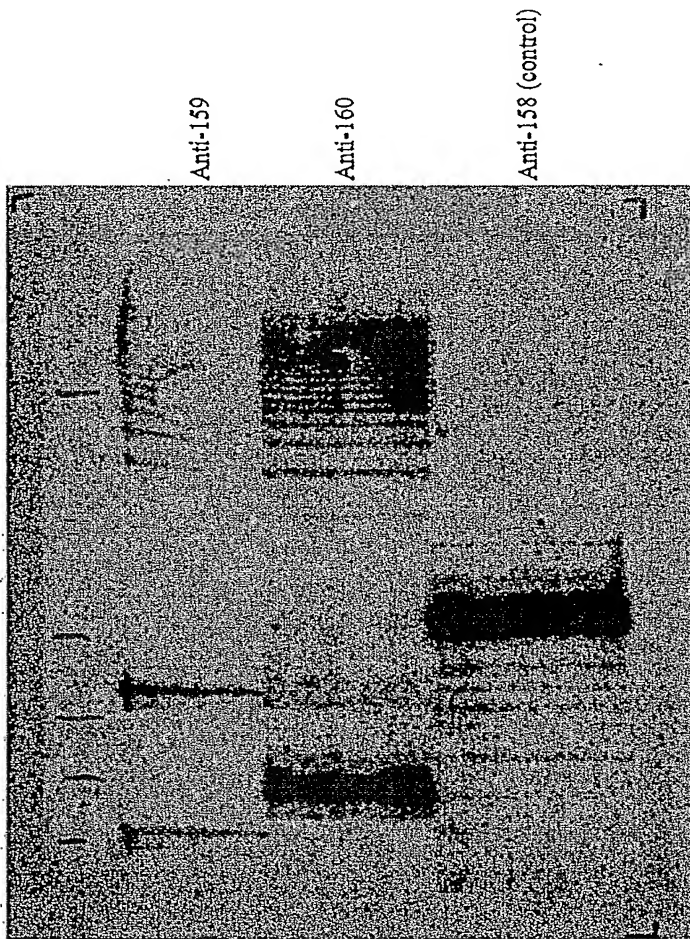




**Note:** 0102 corresponds to SpyM3\_0102 in MGAS315 (M3). This protein is pretty much identical to the M12 19224139 -see also gel "WB M12 (23-11-4b2bis)"-, the M18 spyM18\_0130 and the M5 orf 82. Here the antiserum raised against 0102 recognizes high-weight molecular structures in protein extracts from M12 (strain isolate 2728.)

Figure 109

Western blot on fraction enriched in surface proteins of M6 (2724)

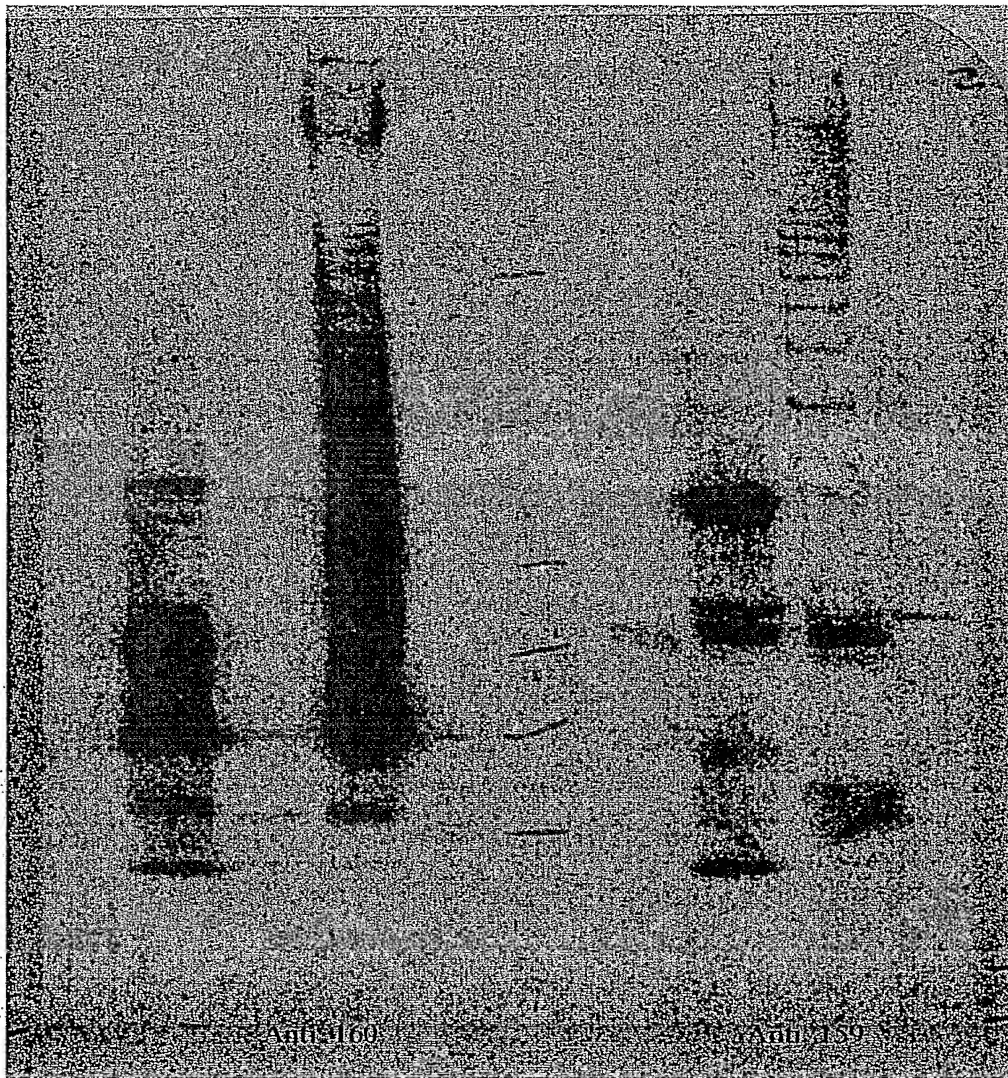


M6 strain isolate 2724

Figure 110

PCT/US05/27239

0160	0159	M6 3650 surf prot	M1 SF370 surf prot	Marker	0160	0159	M6 3650 surf prot	M1 SF370 surf prot
------	------	-------------------	--------------------	--------	------	------	-------------------	--------------------



M6 strain isolate 3650

Figure 111

PCT/US05/27239

M6 2724 fraz art

M6 2724 tot

0160

Marker

158

0160

0159

M6 2724 surf prot

M6 2724 tot

M1 SF370 surf prot

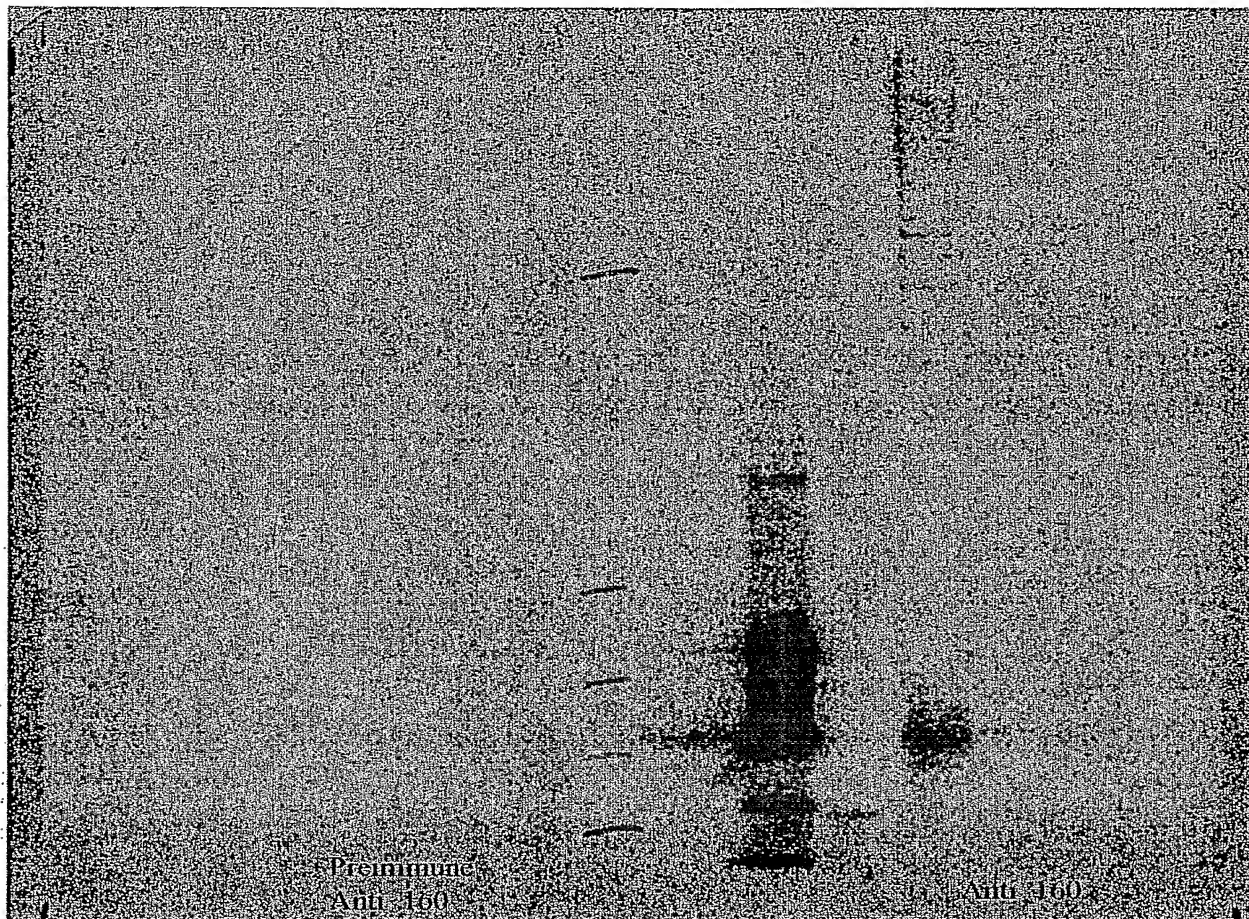


Figure 112



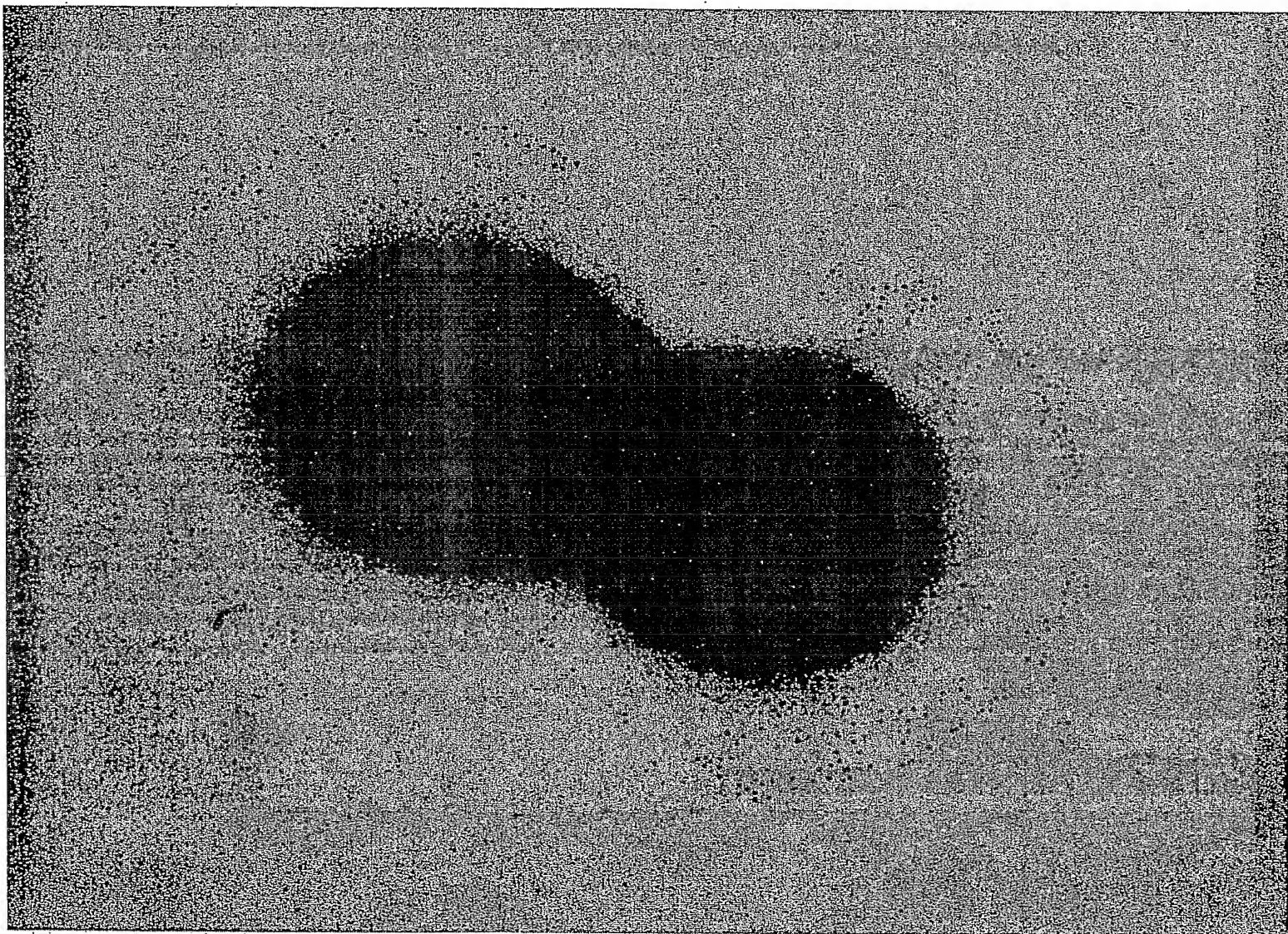


FIGURE 113

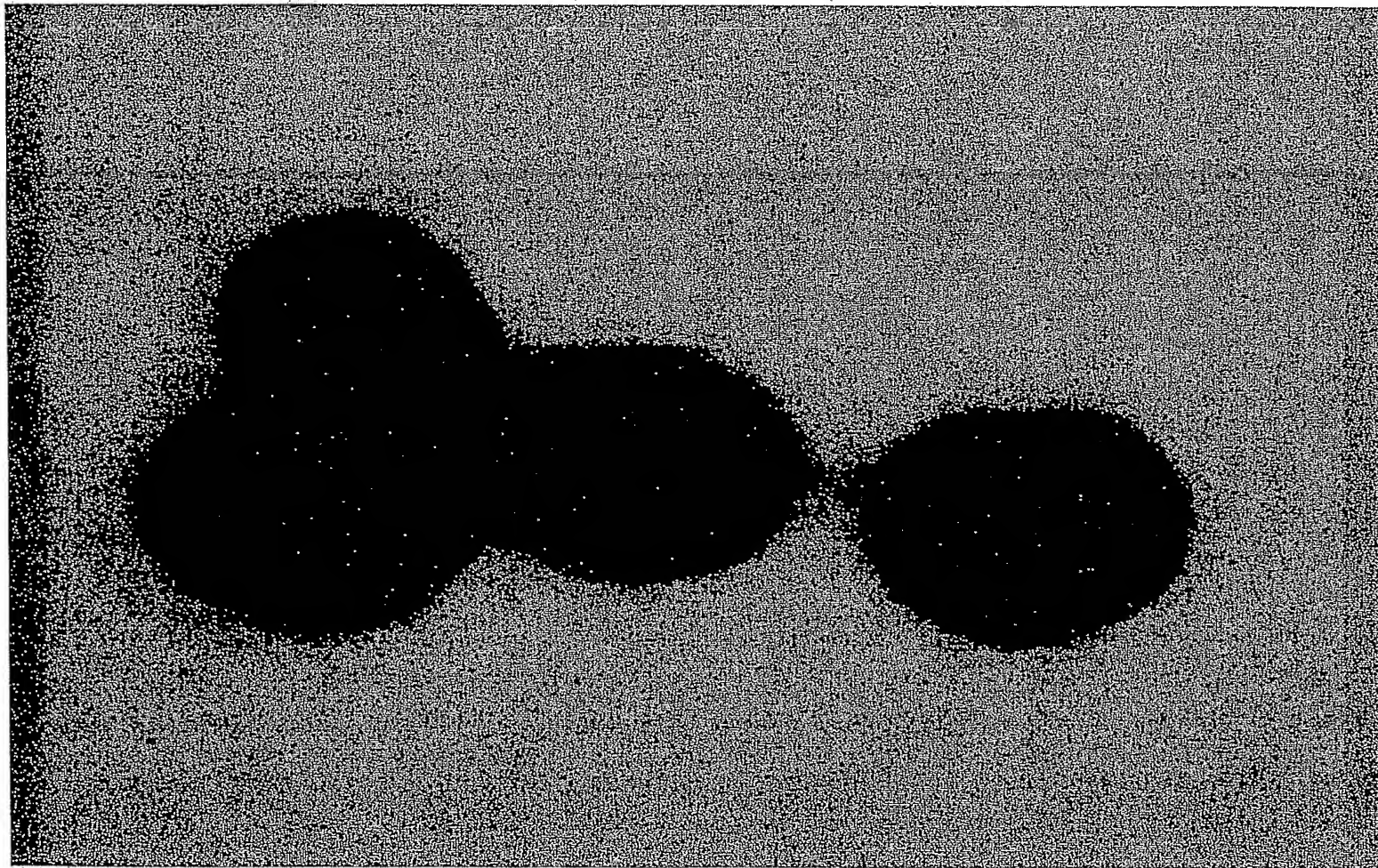


FIGURE 114



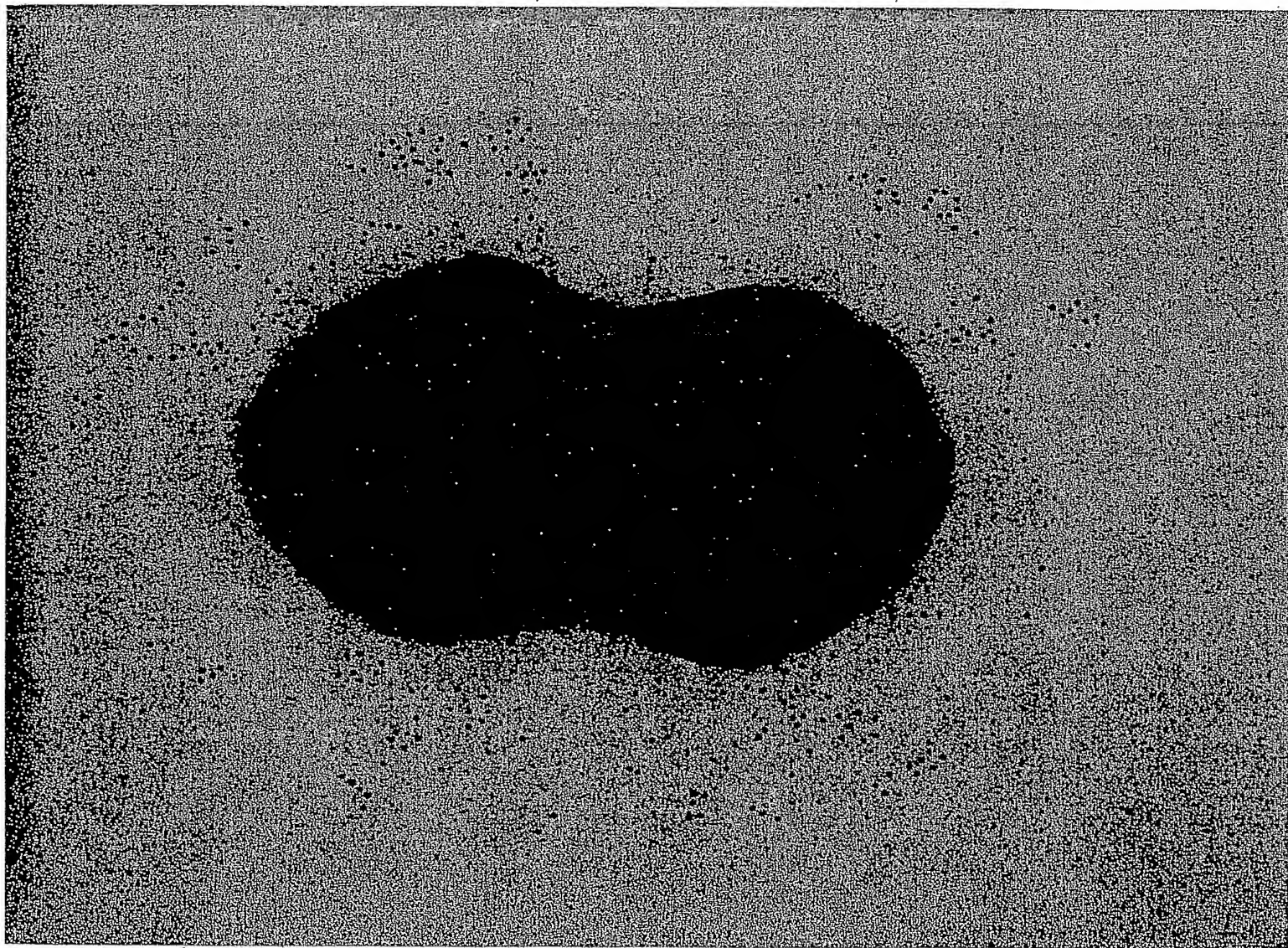


FIGURE 115

PCT/US05/27239 307/487

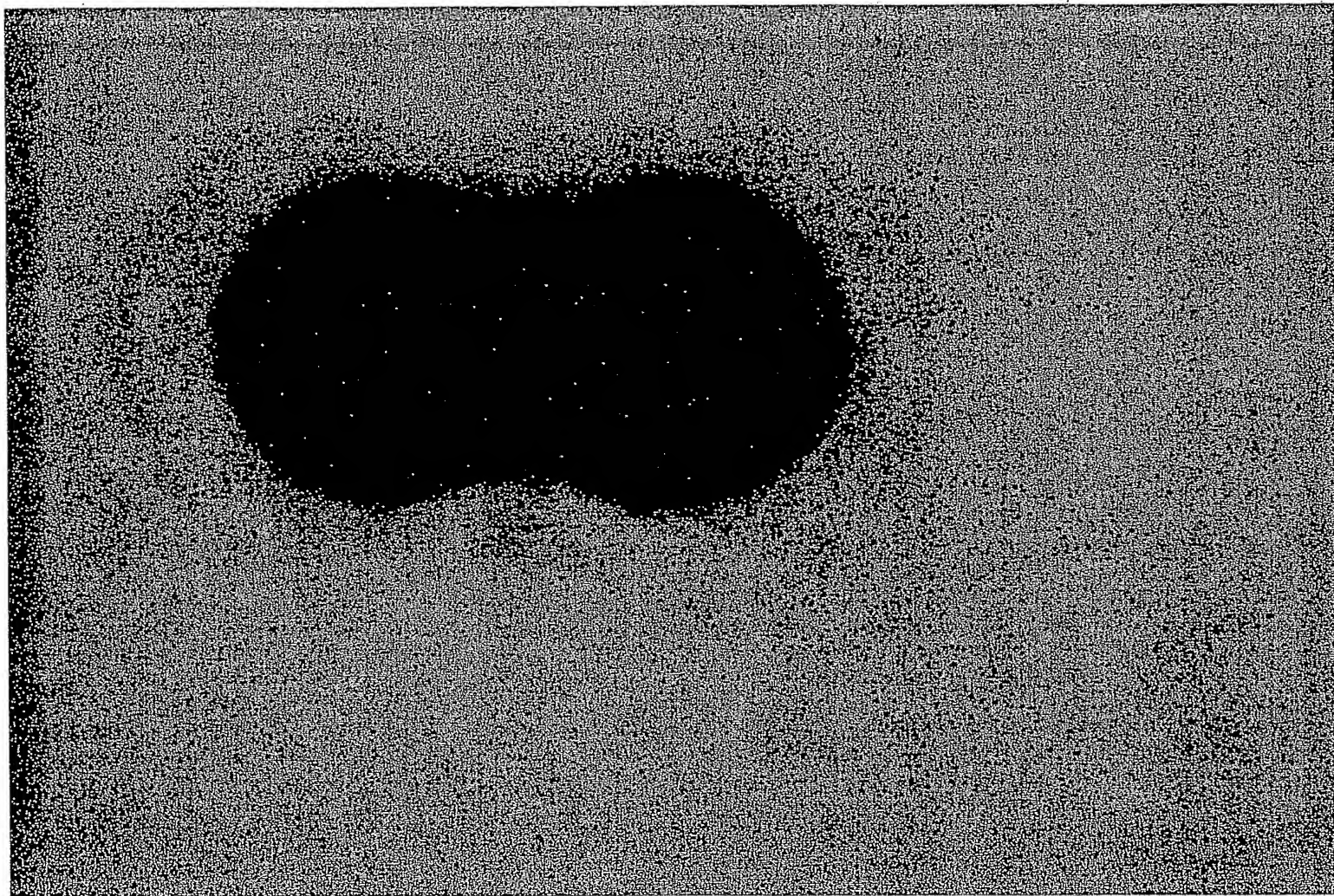


FIGURE 116



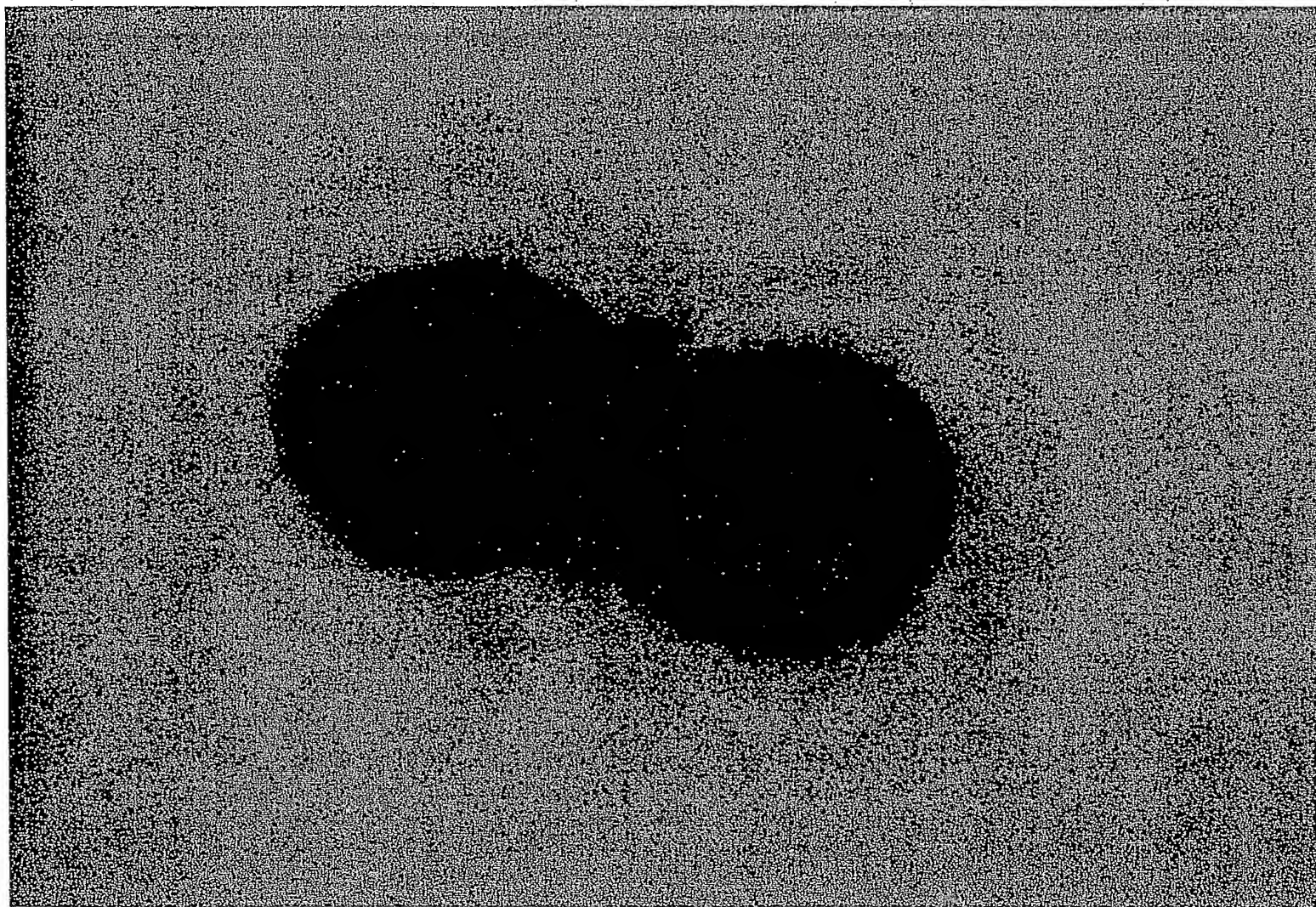


FIGURE 117

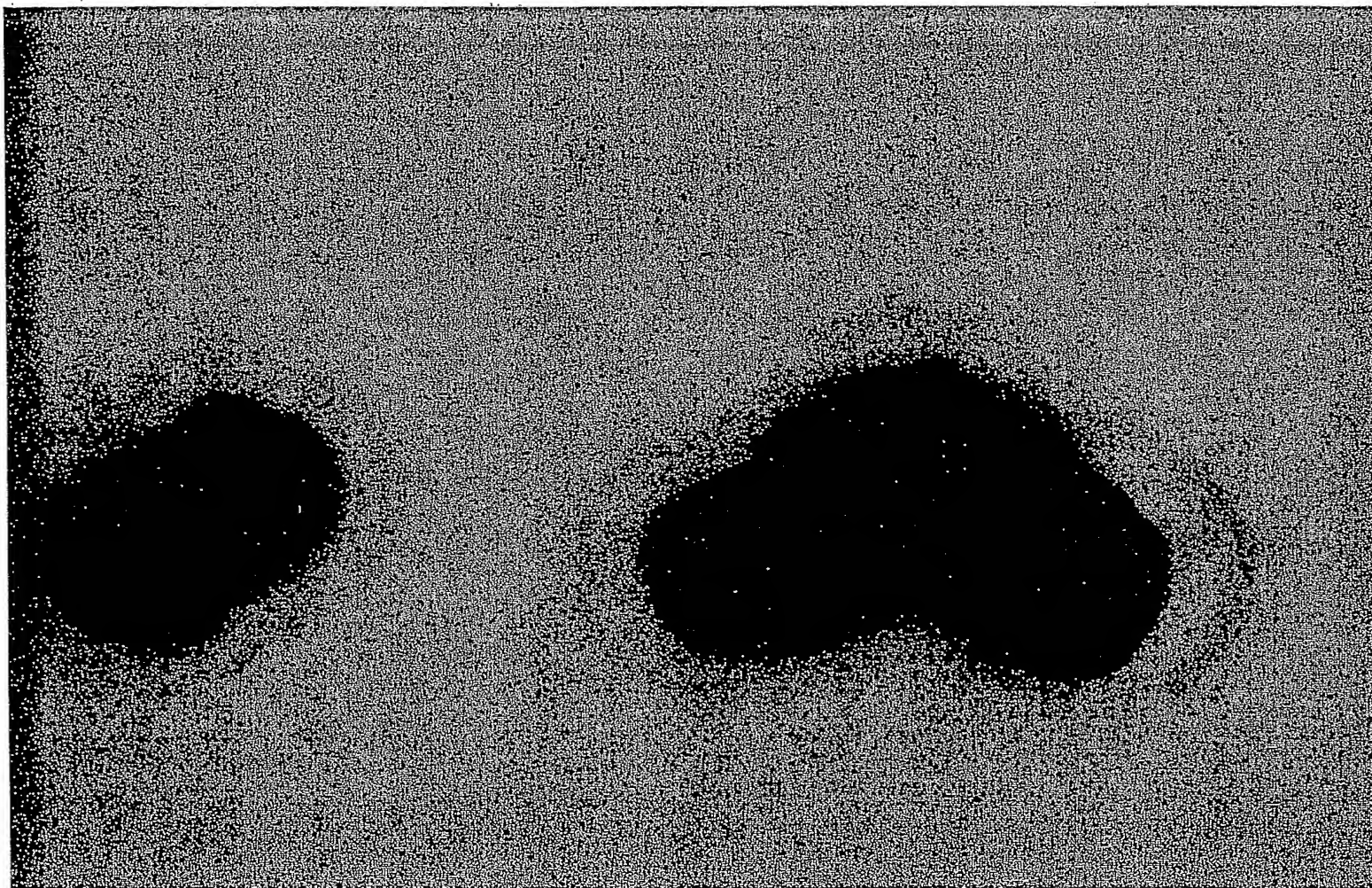


FIGURE 118



PCT/US05/27239

310/487

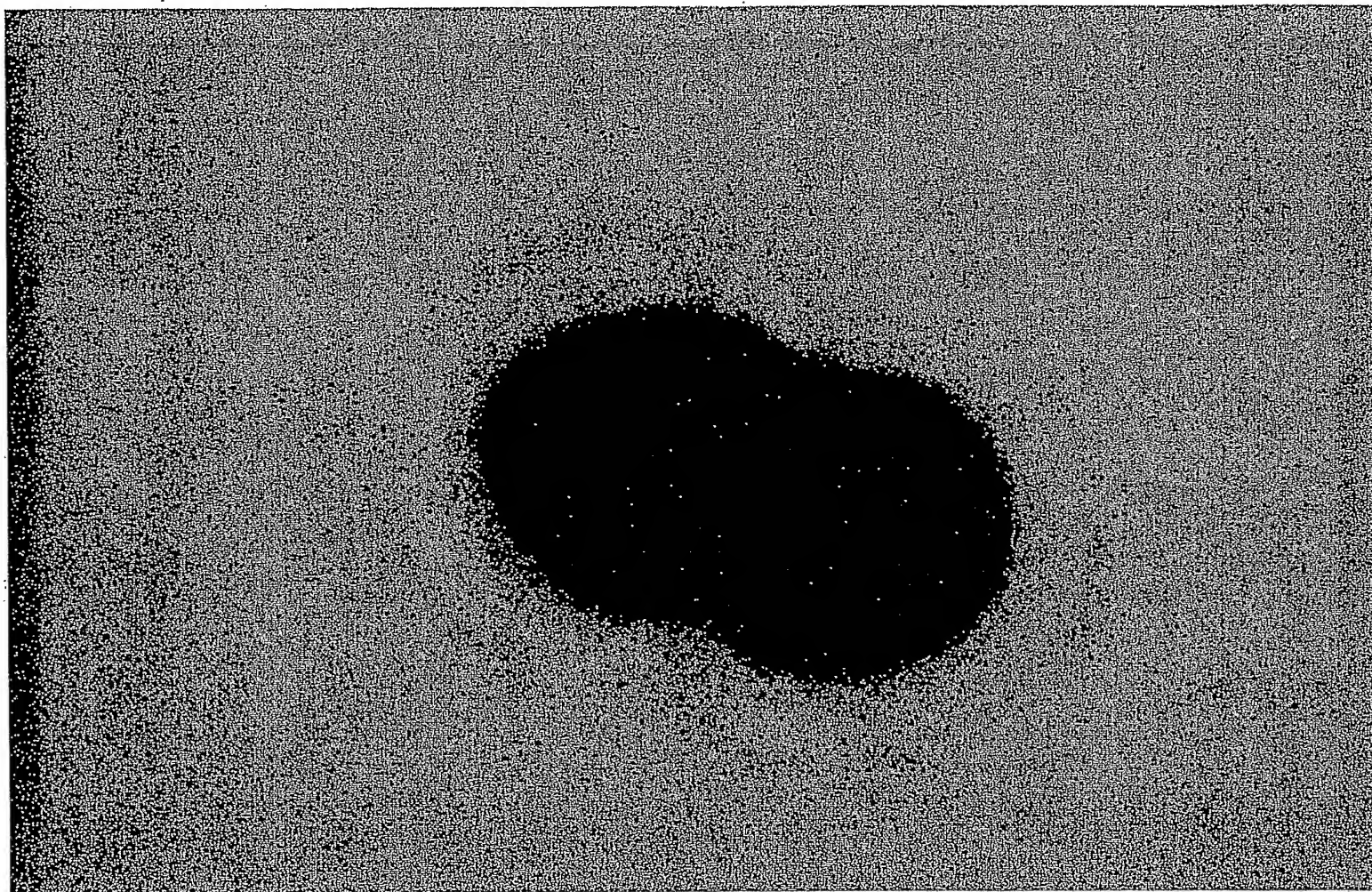


FIGURE 119

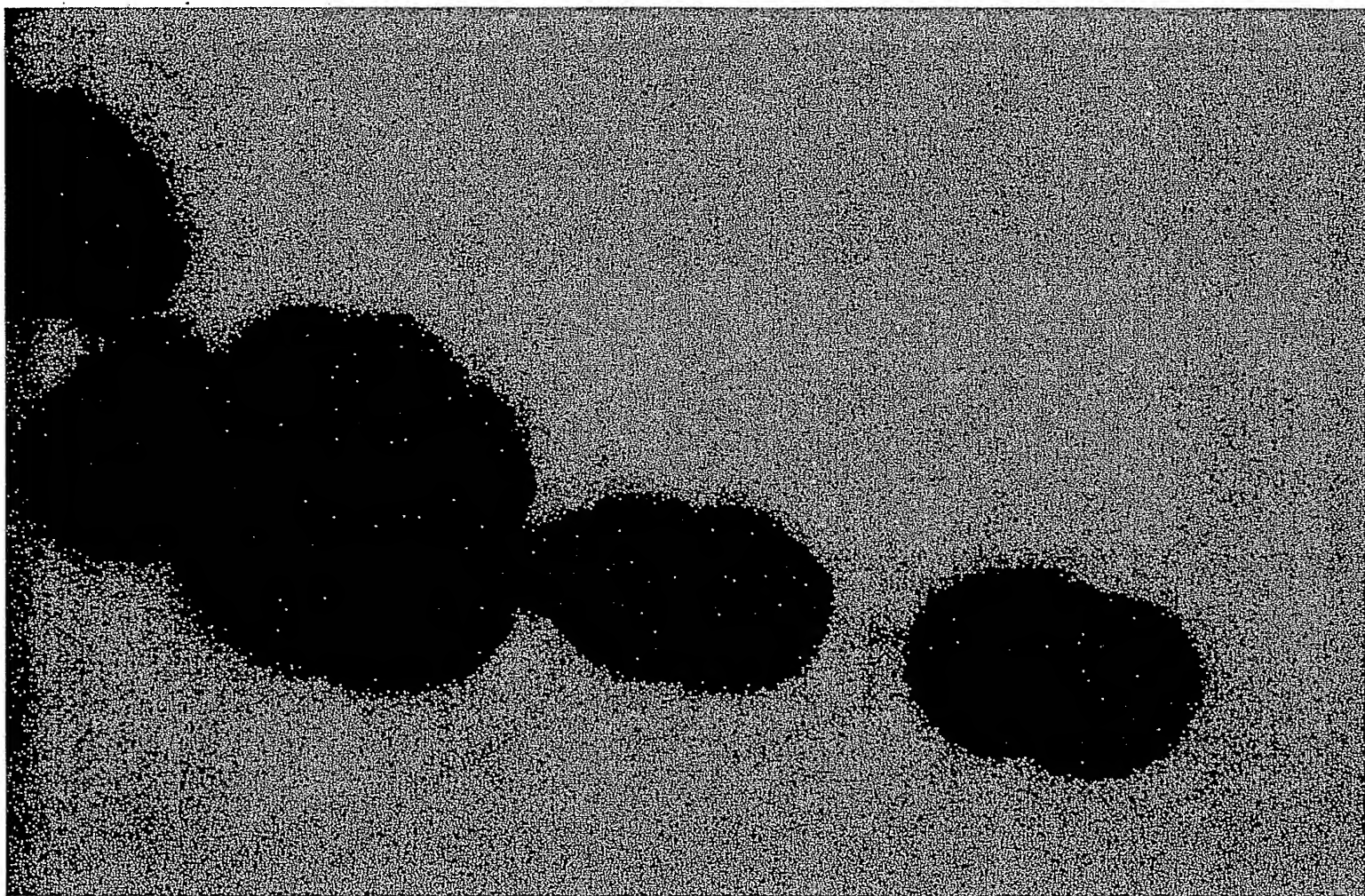


FIGURE 120

312/487

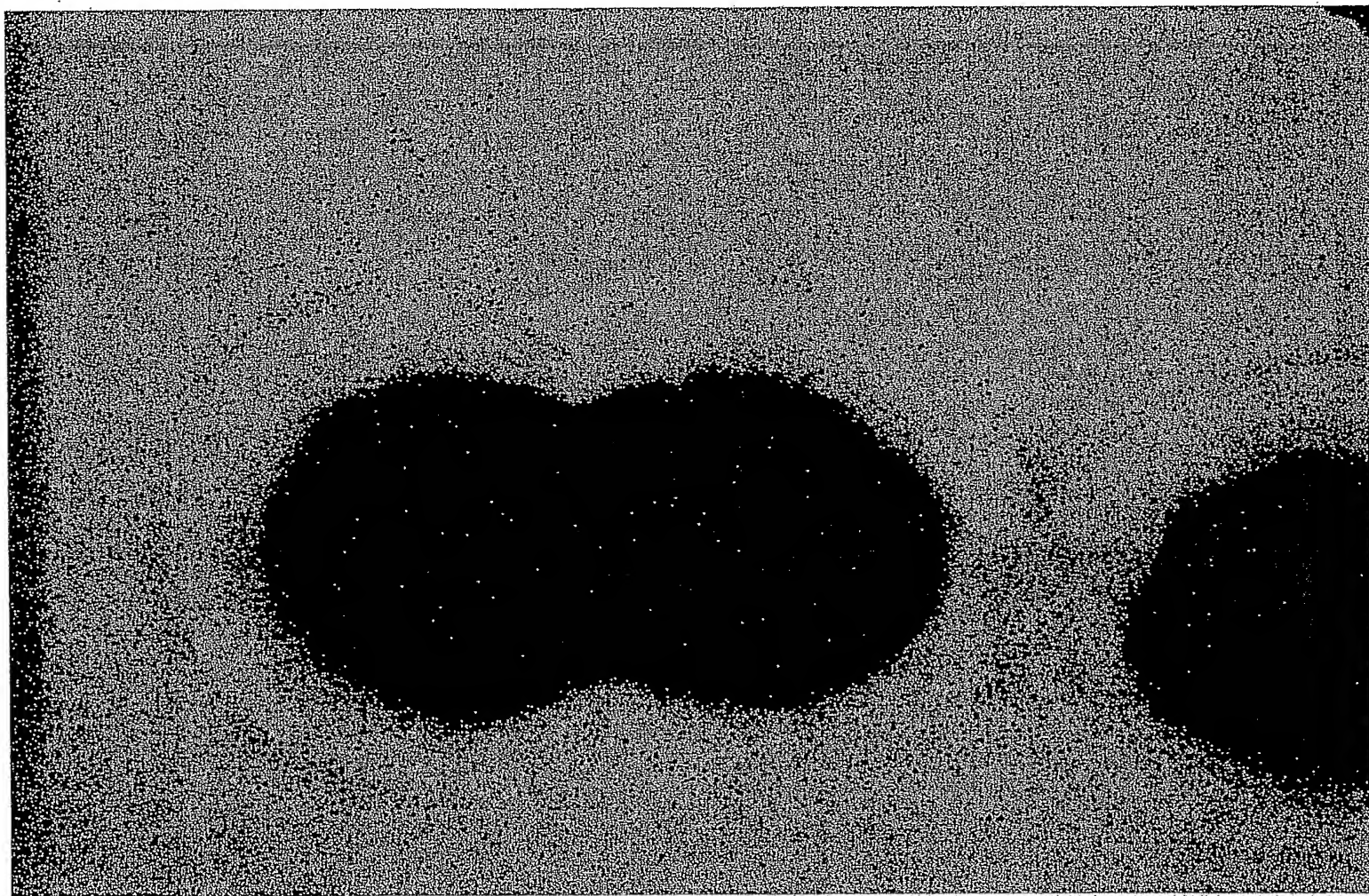


FIGURE 121



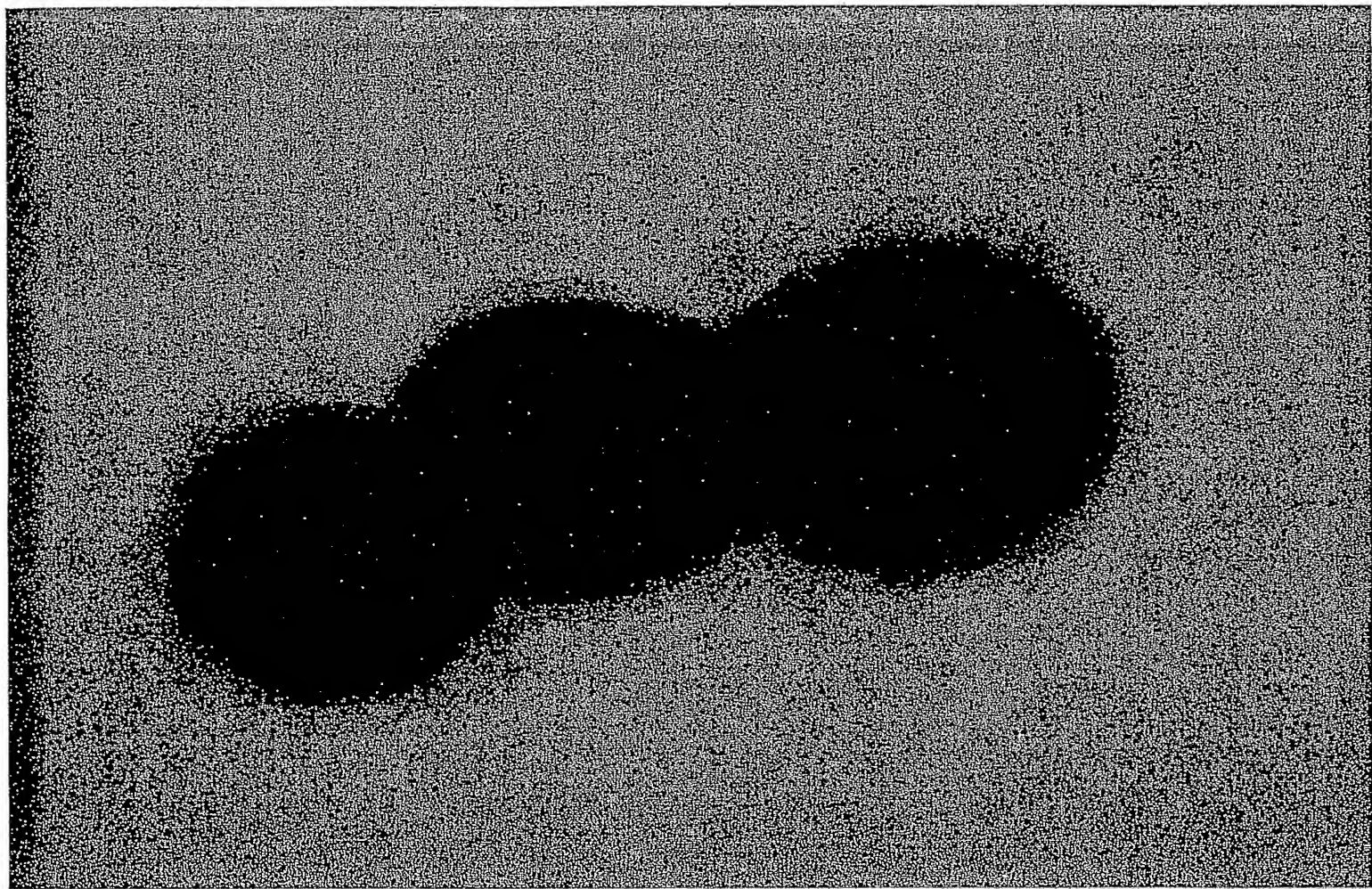


FIGURE 122

314/487

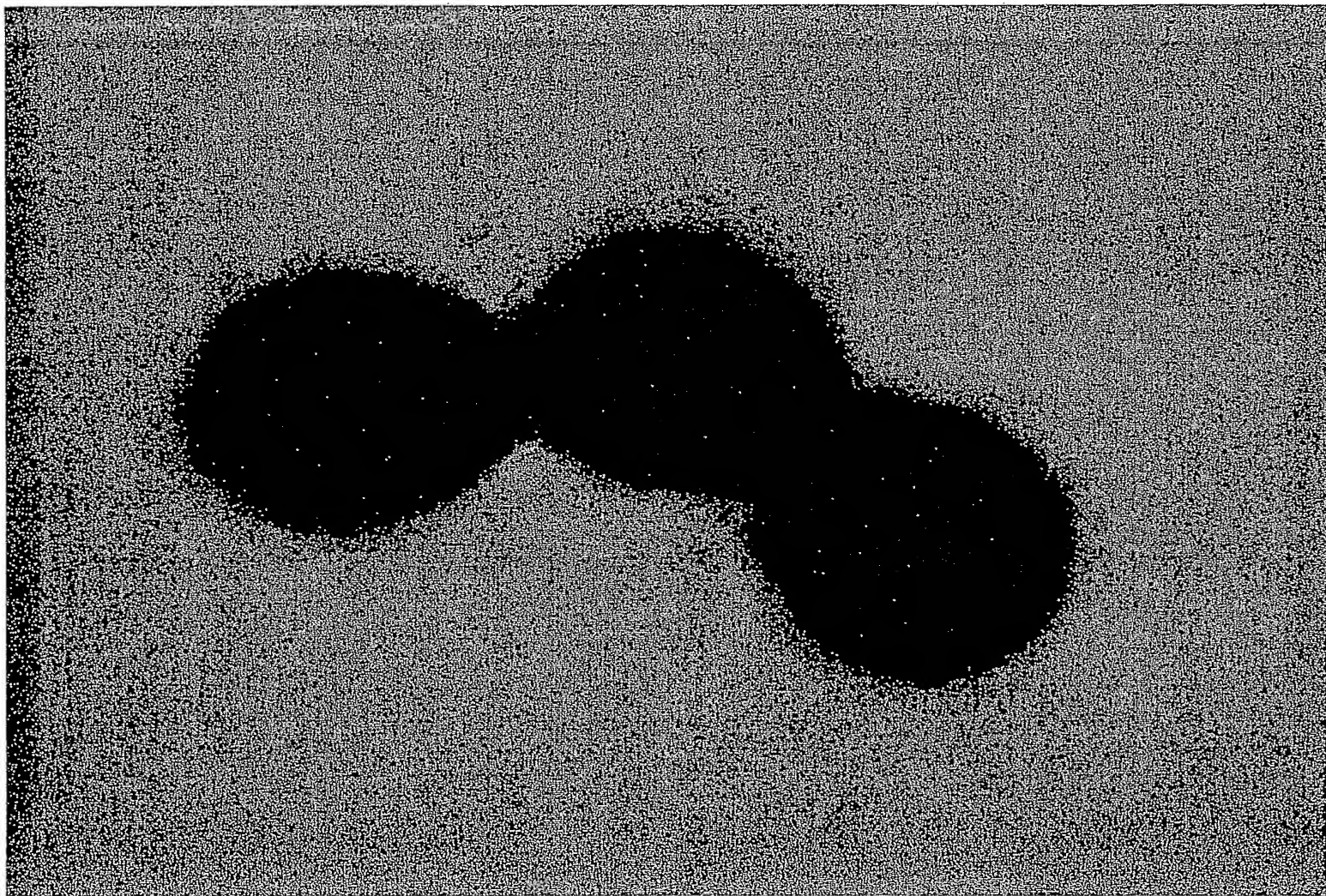


FIGURE 123

315/487

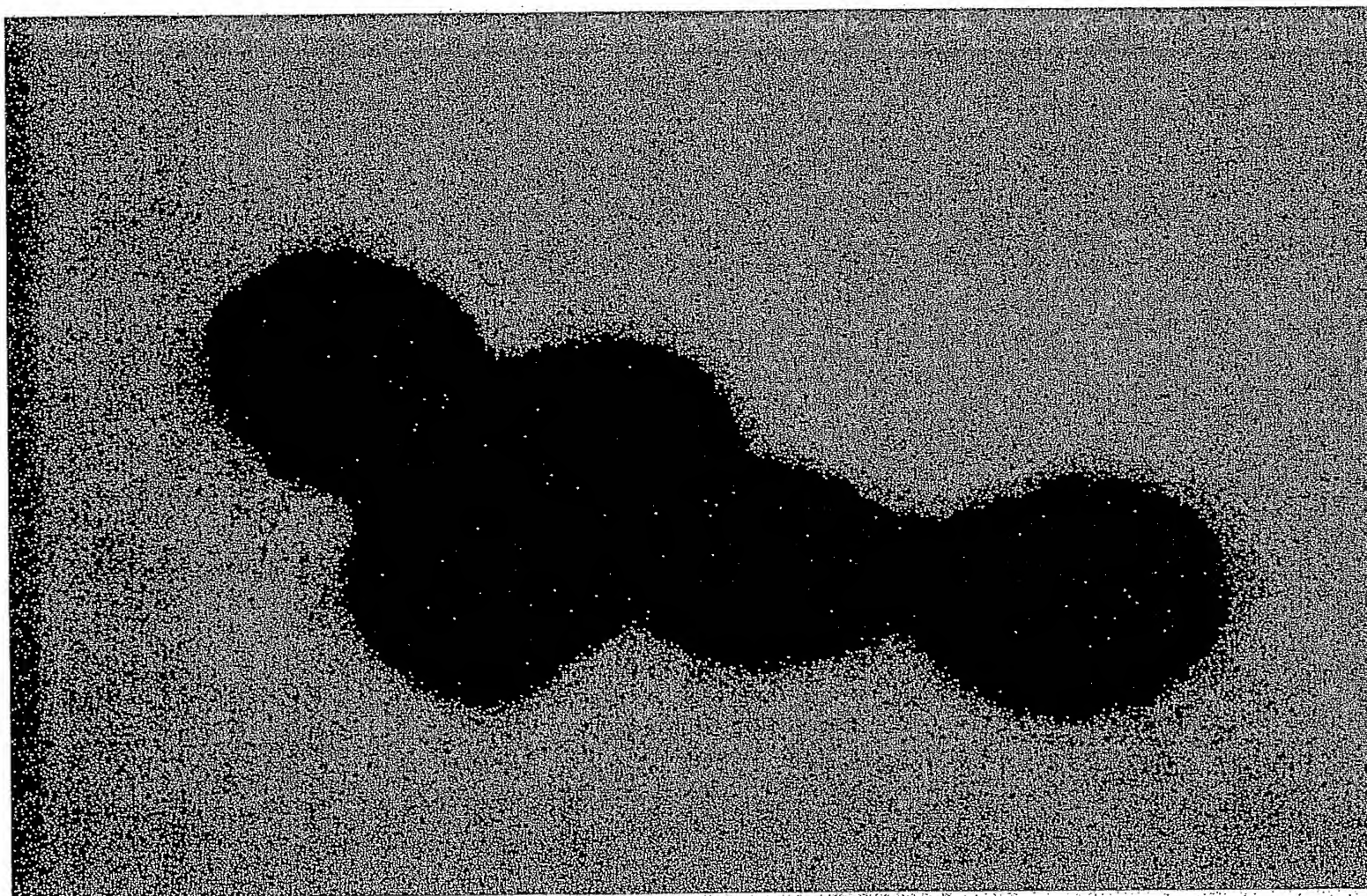


FIGURE 124



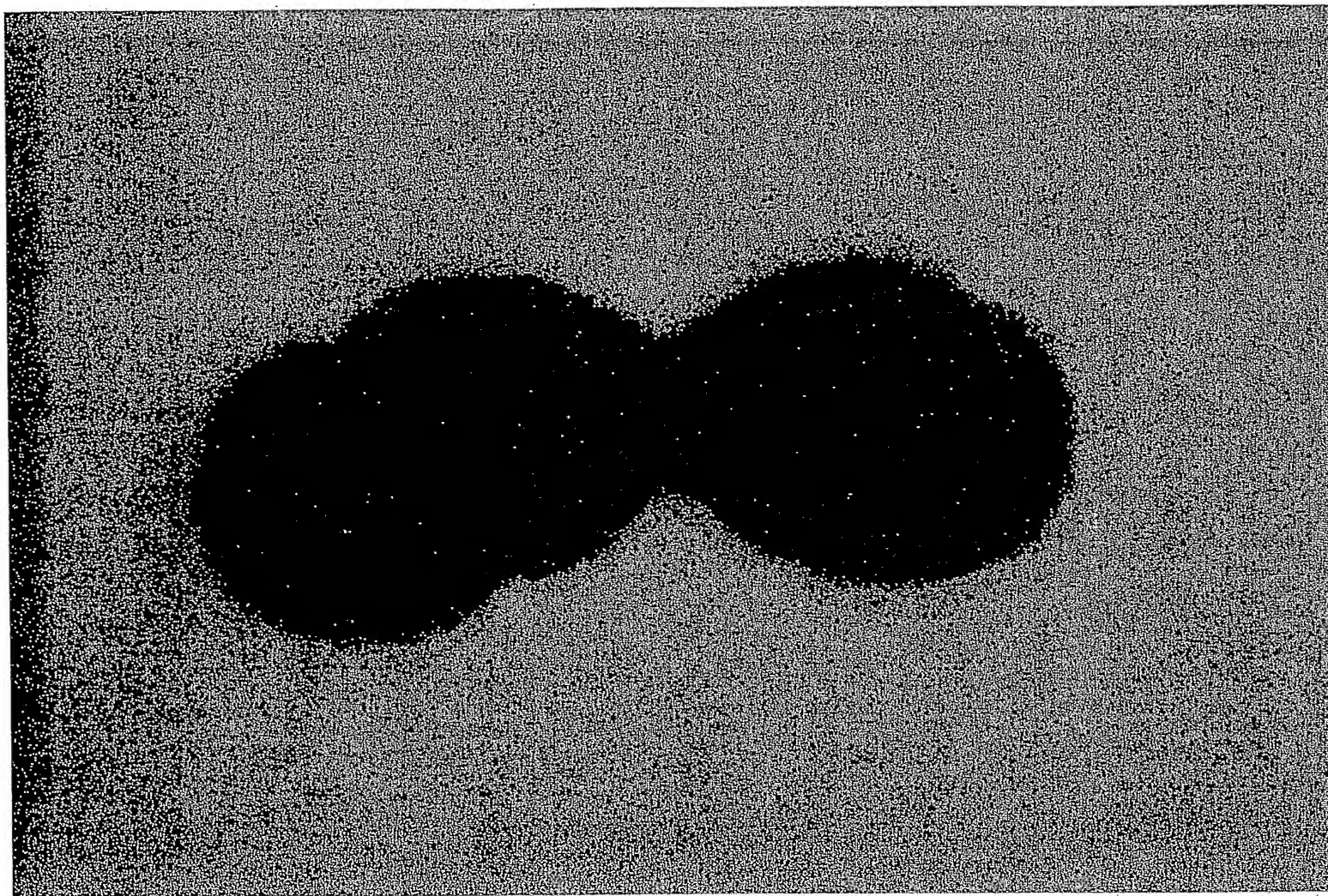


FIGURE 125

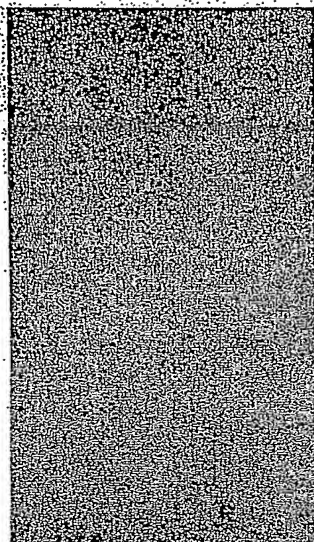
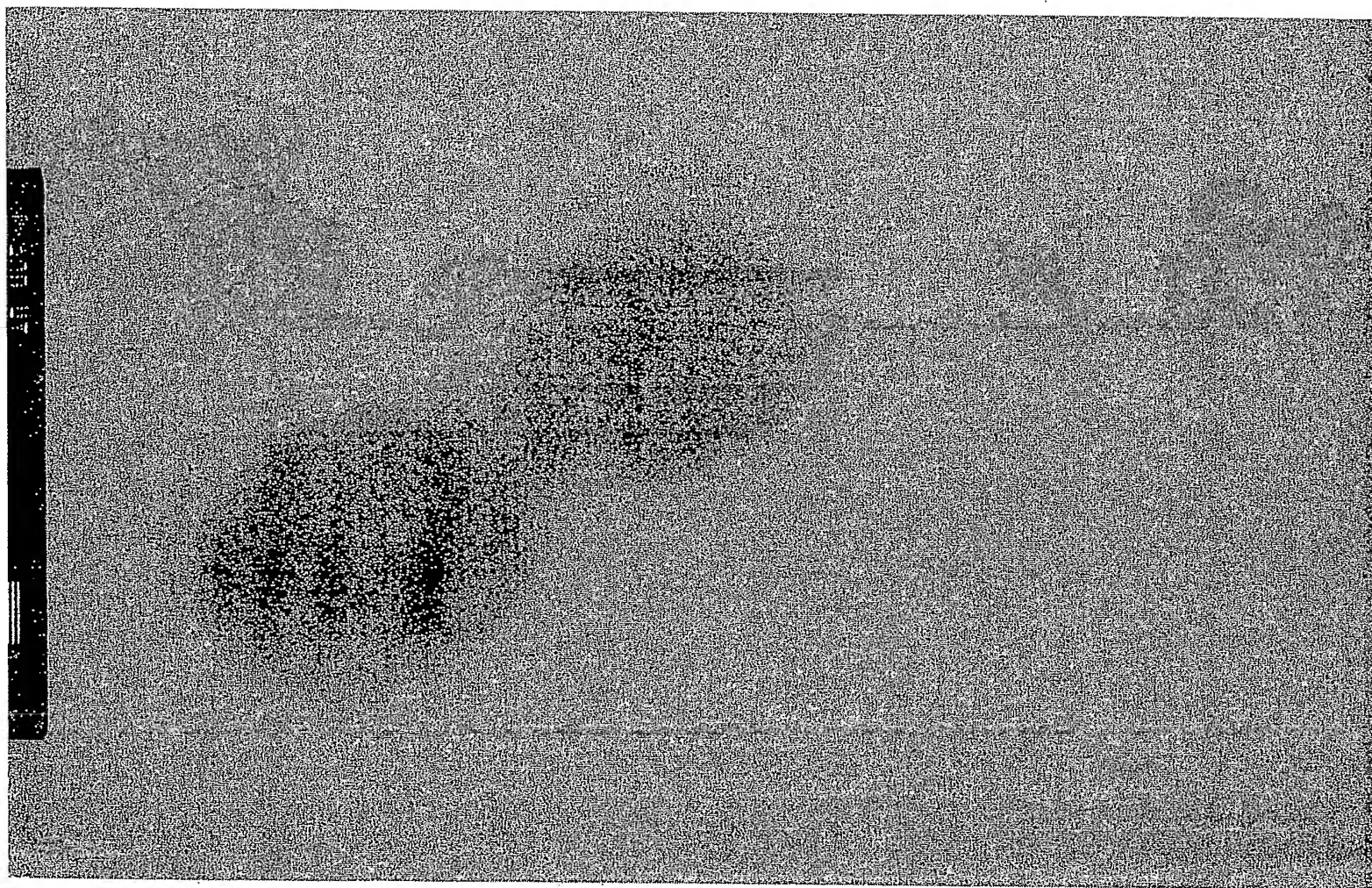


FIGURE 126

Figure 127





PCT/US05/27239 319/487

Figure 128

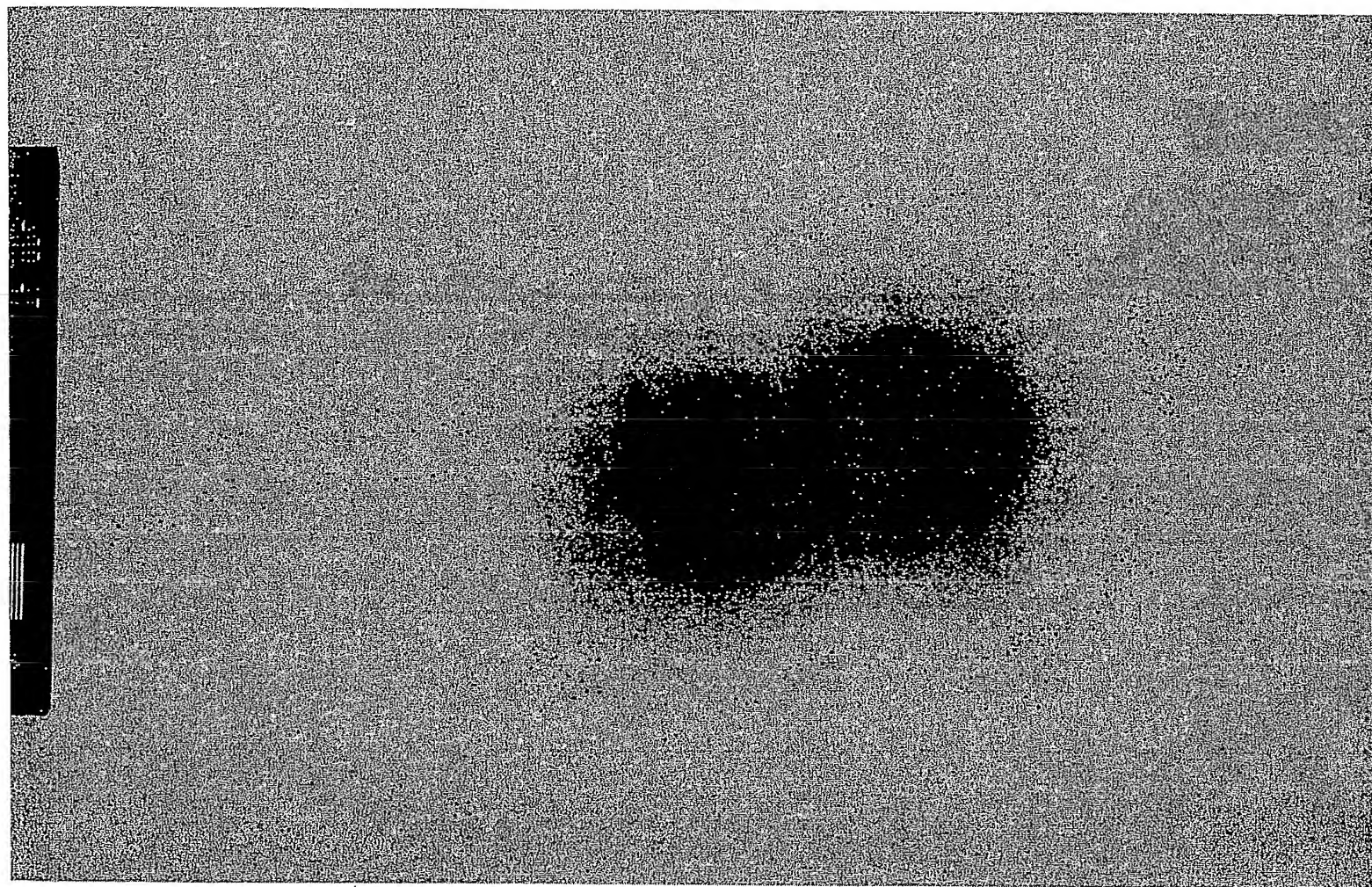


Figure 129

